HAL
archives-ouvertes.fr

DNA Repair in Deinococcus radiodurans

Joanna Timmins

▶ **To cite this version:**

Joanna Timmins. DNA Repair in Deinococcus radiodurans. Biologie structurale [q-bio.BM]. Université Joseph Fourier, 2012. <tel-01301831>

HAL Id: tel-01301831

http://hal.univ-grenoble-alpes.fr/tel-01301831

Submitted on 13 Apr 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

HABILITATION À DIRIGER DES RECHERCHES

Présentée devant l'Université Joseph Fourier, Grenoble

par

Joanna TIMMINS

Chercheur à l'Institut de Biologie Structurale (UMR5075)

le 5 décembre 2012

Jury

Président

RUIGROK Rob, Professeur, Université Joseph Fourier, Grenoble

Rapporteurs:

DE GROOT Nicolaas, Chercheur, CEA, Cadarache

HOPFNER Karl-Peter, Professeur, Gene Zentrum, Munich

TERRADOT Laurent, Chargé de recherche CNRS, IBCP, Lyon

Examinateur:

HANS Fabienne, Maître de Conférences, Université Joseph Fourier, Grenoble

Table of Contents

Résumé

Mon cursus scientifique a débuté avec mes études universitaires en Biochimie à l'Université de Cambridge en Angleterre, suivies de ma thèse effectuée au sein du laboratoire de Winfried Weissenhorn à l'EMBL et de mon travail de post doctorante puis de scientifique à l'ESRF dans le groupe de Biologie Structurale dirigée par Sean McSweeney. Ces travaux m'ont mené jusqu'à l'Institut de Biologie Structurale (IBS) que j'ai rejoins en 2011 avec l'aide d'une bourse ATIP-AVENIR pour y monter ma propre équipe. Durant ces années, j'ai eu l'occasion d'encadrer un grand nombre d'étudiants pour des périodes variant de quelques semaines (dans le cas de stagiaires) à quelques années (dans le cas de thésards), mais aussi de publier de nombreux articles dans des revues à comité de lecture.

En rejoignant l'ESRF en 2003 pour participer à un projet européen (FP6 ; EURAMAN) sur l'étude des protéines lysosomales, je me suis rapidement impliquée dans le projet de génomique structurale centré autour de la bactérie *Deinococcus radiodurans*. Cette bactérie est une bactérie polyextrêmophile et l'un des organismes les plus radiorésistants connu au monde. Cette bactérie présente une résistance impressionnante, notamment aux UV, aux rayonnements ionisants, au peroxyde d'hydrogène, aux températures extrêmes, au dessèchement, au froid et à la famine. Cette capacité de résistance est due à sa structure cellulaire particulière et à son système très perfectionné de réparation de l'ADN, qui lui permet même de reconstituer un génome intact quelques heures à peine après avoir été irradiée avec des doses de rayonnement qui sont létales pour tout autre organisme sur terre.

Avant de détailler mes travaux dans le cadre du projet de génomique structurale et sur les protéines de réparation de l'ADN de *D. radiodurans*, je décris dans un premier temps mon intérêt et mon rôle dans l'étude de la résistance de cette bactérie aux rayons X. En 2006, nous avons en effet irradié *D. radiodurans* avec des doses croissantes de rayons X (expériences effectuées sur la ligne de lumière médicale ID17 de l'ESRF) afin de suivre le taux de survie et l'effet de fortes doses de rayons X sur la courbe de croissance de cette bactérie. Par ailleurs, j'ai également collaboré avec une équipe de chercheurs sur la ligne ID10 de l'ESRF pour le développement de l'imagerie par diffraction cohérente, une étude qui a permis d'obtenir une image reconstituée de *D. radiodurans* à une résolution de près de 30nm, une première sur un si petit échantillon biologique.

Le projet de génomique structurale a débuté en 2002 et s'est poursuivi jusqu'en 2006-2007. Pendant cette période, le groupe de biologie structurale de l'ESRF a sélectionné 100 cibles d'intérêt impliquées dans la résistance aux rayonnements, a réussi à purifier un très grand nombre de ces protéines et a déterminé les structures cristallines de 13 d'entre elles. Je décris donc mon rôle dans l'avancement et la coordination de ce projet à grande échelle. En 2006 je suis nommée coordinatrice du projet de recherche sur *D. radiodurans*.

Ma participation au projet de génomique structurale a débuté avec l'étude d'une enzyme, maltooligosyltrehalose trehalohydrolase, impliquée dans la synthèse d'un disaccharide nommé trehalose jouant un rôle essentiel dans la protection des cellules contre le stress oxydatif. Pour ce projet, j'ai obtenu des cristaux de la protéine intacte et j'ai déterminé sa structure cristalline par la méthode SAD (single wavelength anomalous dispersion) à partir de cristaux de protéine séléniée. Par la suite nous avons déterminé sa structure en complexe avec du sucrose, d'une part et avec du trehalose, d'autre part. Ceci nous a permis d'identifier les résidus catalytiques, le mécanisme permettant la discrimination entre le trehalose et le sucrose et les sites d'interaction avec ces sucres. Ce projet m'a permis de découvrir les différentes techniques utilisées en cristallographie des protéines.

Peu après, j'ai travaillé aux côtés d'Ingar Leiros sur l'étude de la protéine RecO de *D. radiodurans*, une protéine impliquée dans la réparation des cassures double brin de l'ADN. Nous avons déterminé sa structure cristalline, puis étudié sa capacité à interagir avec l'ADN. Nous avons préparé de nombreux mutants afin d'identifier les régions de RecO impliquées dans cette interaction avec l'ADN. Ce premier projet nous a mené à l'étude de l'interaction de RecO avec RecR, une protéine formant un tétramère en forme d'anneau. Très rapidement nous avons obtenu des cristaux de ce complexe, mais ceux-ci ne diffractaient les rayons X que faiblement. Nous avons tout de même réussi à mesurer un jeu de données à une résolution maximale de 3.8 Å et la structure du complexe fut résolue par remplacement moléculaire. Le complexe est formé d'un tétramère de RecR avec une molécule de RecO de part et d'autres de l'anneau de RecR, formant ainsi un hétérohexamère. Là aussi, nous avons préparé de nombreux mutants pour confirmer l'importance des interactions protéines-protéines observées dans notre structure et également pour identifier les résidus de RecO et RecR impliquées dans la reconnaissance de cassures double-brin dans le contexte du complexe. Ces travaux ont ensuite été poursuivis de 2008 à 2011 par un post-doctorant, Jens Radzimanowski, qui a réussi à obtenir une nouvelle forme cristalline de ce complexe et a pu obtenir des données à une résolution de 3.3 Å améliorant ainsi nettement la qualité de nos données structurales sur ce complexe. Par ailleurs, Jens a effectué une étude de SAXS et de SANS (Small-angle X-ray scattering et Small-angle neutron scattering) sur le complexe et sur RecR seule en présence ou non d'ADN nous révélant le mode d'interaction de RecO-RecR avec l'ADN.

De 2008 à 2012, j'ai encadré un étudiant en thèse, Simone Pellegrino, qui a travaillé sur la protéine RecN, jouant un rôle essentiel dans les premières étapes de reconnaissance de cassures double-brin dans l'ADN. Au départ du projet, aucune information structurale n'était disponible pour RecN. Simone a déterminé les structures cristallines de trois fragments de RecN et, avec l'aide d'une enveloppe de la protéine entière déterminée par SAXS, a pu reconstituée une structure quasi atomique de cette protéine très allongée (300 Å de long) appartenant a la famille des SMC (Structural Maintenance of Chromosomes). Simone a également étudié le rôle de la fixation et de l'hydrolyse de l'ATP dans l'activité et l'oligomérisation de RecN.

Dès 2005, je me suis intéressée aux protéines UvrA, UvrB, UvrC et UvrD impliquées dans la réparation des lésions de l'ADN induites par les rayonnements ultra-violets (UV). Contrairement à la majorité des bactéries, *D. radiodurans*, bien connue pour sa très forte résistance à la sècheresse, aux stress oxydants et aux rayonnements UV et ionisants, possède deux formes d'UvrA : UvrA1 et UvrA2. UvrA1 est présente chez tous les procaryotes et joue un rôle essentiel dans la reconnaissance des lésions telles que les dimères de cycobutane. Le rôle d'UvrA2 n'a pas encore été élucidé. Nous avons déterminé la structure cristalline d'UvrA2 en complexe avec de l'ADP à une résolution de 2.3 Å par la méthode MAD (multi-wavelength anomolous dispersion) grâce à la présence de 2 atomes de zinc associés avec chaque monomère d'UvrA2. La structure révèle qu'UvrA2 forme un dimère et que chaque unité est constituée de 2 domaines de fixation de nucléotides (NBD, nucléotide binding domains) repliés l'un sur l'autre formant ainsi deux sites de fixation de l'ATP par monomère. Nos études complémentaires (interaction avec l'ADN lésé et mutagénèse) ont identifié les résidus et domaines impliqués dans la reconnaissance de lésions et nous ont permis de proposer un modèle de l'interaction d'UvrA2 avec l'ADN endommagé.

Plus récemment, nous avons déterminé les structures cristallines d'UvrD, une ADN hélicase de type SF1A (Superfamily 1A) de *D. radiodurans* (drUvrD) en complexe avec différents ADNs et nucléotides, nous offrant plusieurs vues du processus d'ouverture de la double hélice. Chez *D. radiodurans* UvrD joue un rôle essentiel dans de nombreuses voies de

réparation de l'ADN. Nos structures mettent en évidence des changements conformationnels de grandes ampleurs et améliorent notre compréhension des mécanismes moléculaires qui régulent les hélicases de type SF1A. A notre grande surprise, nos données biochimiques ont également démontré que drUvrD déroule l'ADN de 5' vers 3', une polarité caractéristique des hélicases de type SF1B et non SF1A. Cette polarité inversée est obtenue en interagissant avec l'ADN dans le sens opposé et souligne plusieurs différences structurales et fonctionnelles entre les hélicases de type SF1A et SF1B.

En 2005, une post-doctorante, Elin Moe, a visité l'ESRF pendant quelques mois pour travailler sur la détermination des structures des enzymes impliquées dans la voie de réparation par excision de bases chez *D. radiodurans*. Petit à petit, notre intérêt commun pour les mécanismes de réparation de l'ADN nous a rapproché et je me suis impliquée dans ce projet également. En 2007 nous avons obtenu un financement pour un(e) thésard(e) qui partagerait son temps entre l'ESRF et l'université de Tromsø en Norvège. Aili Sarre a donc été recrutée en 2008 pour étudier les trois Endonucléases III de *D. radiodurans*. Les endonucléases III reconnaissent les pyrimidines oxydées. Aili va soutenir sa thèse début 2013. Elle a déterminé les structures cristallines des Endonucléases III-1 et III-3 et a étudié leurs activités respectives et leurs différentes spécificités de substrats.

Le projet de recherche que je souhaite maintenant développer à l'IBS avec l'aide de mon ATIP découle logiquement de ces travaux effectués durant mes 8 années passées à l'ESRF. Mon équipe s'intéresse aux mécanismes moléculaires impliqués dans la reconnaissance et la réparation des lésions de l'ADN chez l'homme et chez *D. radiodurans*. Nos travaux se concentrent sur deux aspects qui sont : *(i)* La dynamique des processus de réparation de l'ADN et *(ii)* La reconnaissance des lésions dans l'ADN. L'objectif est d'utiliser une combinaison de méthodes de Biologie Structurale, d'outils biophysiques et biochimiques et de techniques d'imagerie de fluorescence sur molécules uniques afin de décrypter les processus moléculaires complexes menant à la réparation des lésions de l'ADN.

Summary

Deinococcus radiodurans is an unusual bacterium, displaying an outstanding resistance to desiccation, ionising radiation and DNA damaging agents. I became interested in this organism initially in the context of an ongoing structural genomics project at ESRF, which I joined in 2003. This structural genomics project led to the determination of a dozen crystal structures of proteins associated with the radiation-resistance phenotype of *D. radiodurans*. In addition, a number of complementary approaches were used to study this organism such as irradiation of cells on the medical beamline or coherent diffraction imaging of frozen cells. By 2006, I had become particularly involved in the structural and biochemical studies of several essential DNA repair proteins involved in homologous recombination and the nucleotide-excision repair pathways. A strong collaboration with NorStruct in Tromsø also led me into the base-excision repair field. My interest in the DNA repair mechanisms of *D. radiodurans* brought me to the IBS in 2011 with an ATIP-Avenir grant, where I am now establishing my own team, focusing on the early steps of DNA repair, involving the specific recognition of DNA damages. My objectives are to determine the principles of damage localisation and recognition in cells by using high- and low-resolution structural studies combined with super-resolution imaging of the repair machinery in cells.

Part I: Summary of research career

A. Curriculum vitae

Joanna TIMMINS

Institut de Biologie Structurale
Virus Infection and Cancer Group
41, rue Jules Horowitz
38027 Grenoble Cedex 1

Phone: +33 438 782 209
Mobile: +33 671 680 296
E-mail: Joanna.timmins@ibs.fr

Married, 2 children (2 and 4 years old)
Date of birth: 7th August 1977
Nationalities: British/French

Current status

Head of the *DNA Damage & Repair* team at the IBS, Grenoble since May 2011.

Research experience

2006-2011: Scientist in the Structural Biology Group, European Synchrotron Radiation Facility, Grenoble, France.

Responsibilities: Manager of the ESRF biochemical laboratory, leader of the DNA repair project and coordinator of the in-house *Deinococcus radiodurans* structural genomics project.

Research interests: Structural Biology of key proteins and macromolecular assemblies involved in prokaryotic DNA repair pathways.

2003-2006: Post-doctoral fellow in the Structural Biology Group, European Synchrotron Radiation Facility, Grenoble, France.

Research projects: (i) Structural and functional studies of the human GlcNAc-phosphotransferase enzyme, responsible for targeting of proteins to lysosomes (FP6 project; EURAMAN). Initial work on this project was carried out in Tromsø, Norway (one month visit) and the subsequent development of a large-scale overexpression system in mammalian cells and *in vitro* characterisation was achieved in Grenoble.

(ii) High resolution structure determination of the maltooligosyltrehalose trehalohydrolase enzyme from *D. radiodurans* in complex with disaccharides, as part of the ongoing structural genomics initiative in the Structural Biology Group at ESRF.

(iii) Structural and functional studies of key DNA repair proteins from *D. radiodurans* involved in Recombinational Repair (RecO and RecOR) and Nucleotide Excision Repair (UvrA2).

Education

1998-2003: PhD at the European Molecular Biology Laboratory, Grenoble (registered at the Open University, UK). Thesis supervisor: Dr W. Weissenhorn. PhD awarded on 12th

December 2002. Thesis title: Structural and functional studies of human ADAM12 in myoblast fusion and Ebola virus VP40 in assembly.

1995-1998: B.A. degree in Natural Sciences/Biochemistry (1st class) from the University of Cambridge, UK.

Funding

• In 2008, I was awarded an ESRF PhD Fellowship to fund a graduate student for three years to work on "Structural insight into recombinational repair in *Deinococcus radiodurans*".

• Co-applicant on a research grant application to the Norwegian Research Council (project number: 185269/V30), entitled "Functional and structural biology studies of DNA repair proteins from the radiation resistant bacterium *Deinococcus radiodurans"*. The main proposer is Professor Arne Smalås, head of the Norwegian Structural Biology Centre (NorStruct) at the University of Tromsø in Norway. The grant was awarded in October 2007 to support a graduate student for 4 years. The graduate student started her work in September 2008.

• At ESRF, my research was funded by the in-house biological research programme of ESRF that funds instrumentation and technological development of the protein crystallography beamlines as well as structural biology projects.

• My current research is supported by an ATIP-AVENIR Grant (CNRS/INSERM) awarded to me in 2010 to set-up my own research team at the IBS, Grenoble. The project is entitled: " DNA Repair Machinery: *From fundamental research to drug design*". This funding covers a post-doc position for 2 years and running costs for the laboratory (60k€).

• La Ligue contre le Cancer chose to further sponsor my ATIP-AVENIR research project by covering my salary for 3 years, starting in 2011.

• In October 2011, I was awarded a 1-year grant (28k€) by the Département du Science du Vivant of the CEA for a project entitled: 'Structural and Functional studies of Endonuclease III' to initiate a project in collaboration with the Laboratoire Lésions des Acides Nucléiques from the CEA, Grenoble.

• In January 2012, I was awarded a 2-year grant (50k€) by the Association pour la Recherche sur le Cancer (ARC) for a project entitled: 'Structural studies of protein-DNA complexes involved in the recognition and repair of DNA damages'.

Recent Presentations & Conference contributions

• GTBio Protein Crystallography conference at Lille, November 2007. Oral presentation: Towards a better understanding of DNA repair in *D. radiodurans*.

• 8th Winter Research Conference: Oxidative DNA Damage, at Les Houches, January 2009. Oral presentation: Structure-Function studies of DNA repair proteins from *Deinococcus radiodurans*.

• EMBO Meeting in Amsterdam, August 2009. Poster: Structural and mutational studies of *Deinococcus radiodurans* UvrA2 provide insight into DNA binding and damage recognition.

• Invited speaker at the Institut de Biologie Structurale in Grenoble, November 2009. Title: Towards a better understanding of DNA repair in the extreme-radiation resistant bacterium *D. radiodurans.*

• Invited speaker at the Plasticité des Génomes 2010 Meeting in Grenoble, November 2010. Title: Les mécanismes de réparation de l'ADN chez *D. radiodurans.*

• Invited speaker at the Journées Scientifiques de l'IBS, June 2011. Title: Structural Insight into DNA Repair.

• Invited speaker at the Institut de Génétique et Microbiologie, Orsay, October 2012. Title: Études structurales et fonctionnelles des protéines Uvr de *Deinococcus radiodurans*.

• Invited speaker at the Institut de Biologie et Chimie des Protéines, Lyon, December 2012. Title: Études structurales et fonctionnelles des mécanismes de réparation de l'ADN.

Additional scientific activities

• Member of the Partnership for Structural Biology Laboratory Committee from 2006-2011.

• Member of the editorial board of the Partnership for Structural Biology Newsletter as the ESRF representative from 2007 to 2010, and now as the IBS representative.

• Local contact on the MX beamlines (ID14, ID23 & ID29) at the ESRF from 2003 to 2011.

• Member of the CN4 scientific commission of the Fondation ARC that evaluates research proposals, since 2012.

• Reviewer for Current Biology and Biochemistry Journals.

• Reviewer of ANR grant proposals.

• Member of the thesis advisory committee of four PhD students

B. Supervision of research staff

Supervision of trainees and Master's students:

2012 Supervision of Sandrine Gauthier, Master1 student (4 months)

Functional characterisation of human Endonuclease III (hNTH1)

2011 Supervision of Catherine Surr, summer trainee (Licence; 2 months)

Purification of D. radiodurans UvrC fragments

2010 Co-supervision of Virginie Carbonell, Master2 Pro student (6 months)

Structural characterisation of 4-coumarate:coA ligase

2007-2008 Co-supervision of Therese Johansson, Master's student from Sweden (1 year)

Characterisation of D. radiodurans UvrB

2006-2007 Supervision of Sarah Cake, BSc. Degree placement student (1 year)

Characterisation of the Par proteins from D. radiodurans

2005 Supervision of Sofia Caria, Master's student from Portugal (1 year)

Purification and crystallisation of D. radiodurans UvrA2

Supervision of laboratory technicians:

2009-2010 Supervision of Ulrike Kapp, technician

Structural studies of H. pylori DnaB helicase

2007-2011 Supervision of Samira Acajjaoui, technician

Structural and functional studies of D. radiodurans Uvr proteins

Supervision of graduate students:

2010-present Co-supervision of Morgane Lourdin, PhD student, UJF, Grenoble

Characterisation of Deinococcus radiodurans Fpg enzyme

2009-present Co-supervision of Kjertsi Lien, visiting PhD student for 6 months (Norway)

Structural studies of key DNA repair enzymes from D. radiodurans

2008-2012 Co-supervision of Aili Sarre, PhD student (Norway)

Structural and functional studies of D. radiodurans EndoIII enzymes

2008-2012 Supervision of Simone Pellegrino, PhD student. Thesis defended 28th Feb2012.

Structural studies of D. radiodurans RecN protein

Supervision of post-docs:

2011-present Supervision of Meike Stelter, post-doc

Structural studies of protein-DNA complexes involved in DNA repair

2008-2011 Supervision of Tobias Klar, post-doc

Characterisation of the UvrA and UvrB proteins from D. radiodurans

2008-2011 Supervision of Jens Radzimanowski, post-doc

Structural studies of the RecOR complex from D. radiodurans

2008-2011 Supervision of Meike Stelter, post-doc

Structural studies of HpDnaB and DrUvrD helicases

C. Publications

1. Sarre A, Ökvist M, Klar T, Hall D, Smalås AO, McSweeney S, Timmins J and Moe E. Structure/function studies of three Endonuclease III enzymes from the extreme radiation resistant bacterium *Deinococcus radiodurans*. (2012) *Manuscript in preparation*.

2. Radzimanowski J, Dehez F, Round A, Bidon-Chanal A, McSweeney S and Timmins J. New insights into the assembly and DNA binding of the RecOR complex. (2012) *Manuscript in preparation*.

3. Acajjaoui S, Stelter M, McSweeney S and Timmins J. Structural and functional characterization of drUvrD provide new insights into DNA unwinding and helicase polarity. (2012) *Submitted to Plos One*.

4. Pellegrino S, Radzimanowski J, McSweeney S and Timmins J. Crystal structures of *D. radiodurans* RecN: New insight into double-strand break repair. Structure. (2012) *in press*.

5. Lallemand LA, Zubieta C, Lee SG, Wang W, Acajjaoui S, Timmins J, McSweeney S, Jez JM, McCarthy JG and McCarthy AA. A structural basis for the biosynthesis of the major chlorogenic acids found in coffee. Plant physiology. (2012) 160 p.249-260.

6. Moe E, Hall DR, Leiros I, Talstad V, Timmins J, McSweeney S. Structure/function studies of an unusual 3-methyladenine DNA glycosylase II (AlkA) from *Deinococcus radiodurans*. Acta Crys. (2012) D68 p.703-712.

7. Stelter M, Gutsche I, Kapp U, Bajic G, Goret G, Jamin M, Timmins J, Terradot L. Architecture of a dodecameric bacterial replicative helicase. Structure. (2012) 20 p.554-564.

8. Pellegrino S, de Sanctis D, McSweeney S, Timmins J. Expression, purification and preliminary structural analysis of the coiled-coil domain of *Deinococcus radiodurans* RecN. Acta Crys. (2012) F68 p.218-221.

9. Pellegrino S, Radzimanowski J, McSweeney S, Timmins J. Expression, purification and preliminary structural analysis of the head domain of Deinococcus radiodurans RecN. Acta Crys. (2012) F68 p.81-84

10. Lima E, Wiegart L, Pernot P, Howells M, Timmins J, Zontone F, Madsen A. Cryogenic X-ray diffraction microscopy for biological samples. Phys Rev Lett. (2009) 103 (19) p.198102.

11. Timmins J, Gordon E, Caria S, Leonard G, Acajjaoui S, Kuo MS, Monchois V, McSweeney S. Structural and mutational analyses of Deinococcus radiodurans UvrA2 provide insight into DNA binding and damage recognition by UvrAs. Structure. (2009) 17 (4) p.547-58.

12. Timmins J, Leiros I, McSweeney S. Crystal structure and mutational study of RecOR provide insight into its mode of DNA binding. EMBO J. (2007) 26 (13) p.3260-71.

13. Timmins J, McSweeney S. XPB: An essential helicase involved in both transcription and repair of DNA. Mol Cell. (2006) 22 (2) p.149-50.

14. Leiros HK, Timmins J, Ravelli RB, McSweeney SM. Is radiation damage dependent on the dose rate used during macromolecular crystallography data collection? Acta Crys. Section D. (2006) 62 (Pt 2) p.125-32.

15. Garcel A, Gout E, Timmins J, Chroboczek J, Fender P. Protein transduction into human cells by adenovirus dodecahedron using WW domains as universal adaptors. J Gene Med. (2006) 8 (4) p.524-31.

16. Timmins J, Leiros HK, Leonard G, Leiros I, McSweeney S. Crystal structure of maltooligosyltrehalose trehalohydrolase from Deinococcus radiodurans in complex with disaccharides. J. Mol. Biol. (2005) 347 (5) p.949-63.

17. Timmins J, Leiros I, Hall DR, McSweeney S. Crystal structure and DNA-binding analysis of RecO from *Deinococcus radiodurans*. EMBO J. (2005) 24 (5) p.906-18.

18. Hoenen T, Volchkov V, Kolesnikova L, Mittler E, Timmins J, Ottmann M, Reynard O, Becker S, Weissenhorn W. VP40 octamers are essential for Ebola virus replication. J Virol. (2005) 79 (3) p.1898-905.

19. Sola M, Bavro VN, Timmins J, Franz T, Ricard-Blum S, Schoehn G, Ruigrok RW, Paarmann I, Saiyed T, O'Sullivan GA, Schmitt B, Betz H, Weissenhorn W. Structural basis of dynamic glycine receptor clustering by gephyrin. EMBO J. (2004) 23 (13) p.2510-9.

20. Timmins J, Ruigrok RW, Weissenhorn W. Structural studies on the Ebola virus matrix protein VP40 indicate that matrix proteins of enveloped RNA viruses are analogues but not homologues. FEMS Microbiol Lett. (2004) 233 (2) p.179-86.

21. Timmins J, Schoehn G, Kohlhaas C, Klenk HD, Ruigrok RW, Weissenhorn W. Oligomerization and polymerization of the filovirus matrix protein VP40. Virology. (2003) 312 (2) p.359-68.

22. Gomis-Rüth FX, Dessen A, Timmins J, Bracher A, Kolesnikowa L, Becker S, Klenk HD, Weissenhorn W. The matrix protein VP40 from Ebola virus octamerizes into pore-like structures with specific RNA binding properties. Structure. (2003) 11 (4) p.423-33.

23. Timmins J, Schoehn G, Ricard-Blum S, Scianimanico S, Vernet T, Ruigrok RW, Weissenhorn W. Ebola virus matrix protein VP40 interaction with human cellular factors Tsg101 and Nedd4. J Mol Biol. (2003) 326 (2) p.493-502.

24. Timmins J, Scianimanico S, Schoehn G, Weissenhorn W. Vesicular release of ebola virus matrix protein VP40. Virology. (2001) 283 (1) p.1-6.

25. Scianimanico S, Schoehn G, Timmins J, Ruigrok RH, Klenk HD, Weissenhorn W. Membrane association induces a conformational change in the Ebola virus matrix protein. EMBO J. (2000) 19 (24) p.6732-41.

Part II: Past research (2003-2011)

Introduction

My PhD work (1998-2002) was carried out at the European Molecular Biology Laboratory (EMBL) Grenoble Outstation in the group of Winfried Weissenhorn and included two separate studies. The first project investigated the role of human ADAM 12, a member of the growing ADAM (A Disintegrin And Metalloproteinase) protein family, involved in cell-cell and cell-matrix interactions, while the second and principal project of my PhD focused on the study of the matrix protein (VP40) of Ebola virus, a key protein essential for efficient viral assembly and budding. The use of various techniques (X-ray crystallography, electron microscopy, biochemical assays and cell biology) allowed us to study the cellular localisation of VP40 and its role in virus budding, to characterise its ability to oligomerise and to identify cellular binding partners (Figure 1). This work led to several publications in peer-reviewed journals (Scianimanico, Schoehn et al. 2000; Timmins, Scianimanico et al. 2001; Gomis-Roth, Dessen et al. 2003; Timmins, Schoehn et al. 2003a; Timmins, Schoehn et al. 2003b; Hoenen, Volchkov et al. 2005). In 2004, we were invited to write a review entitled: "Structural studies on the Ebola virus matrix protein VP40 indicate that matrix proteins of enveloped RNA viruses are analogues but not homologues" for publication in FEMS Microbiology Letters (Timmins, Ruigrok et al. 2004).

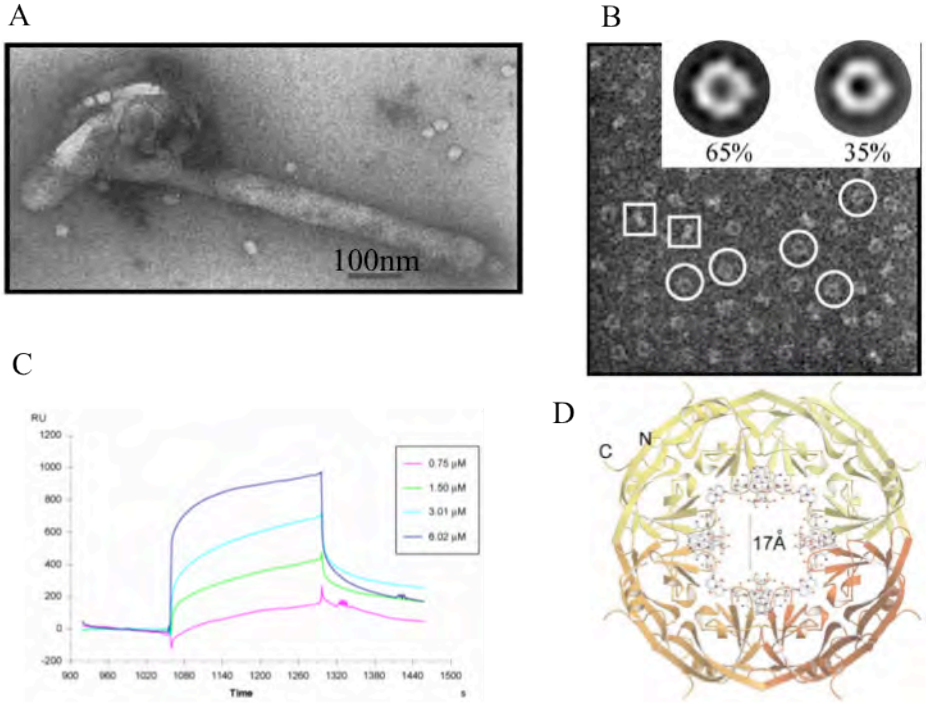


Figure 1: Characterisation of the Ebola virus matrix protein, VP40. (A) Electron micrograph of an isolated virus-like particle secreted by mammalian cells overexpressing VP40. (B) Negative stain electron micrograph of oligomeric VP40 revealing the presence of both hexamers and octamers. (C) Surface Plasmon resonance study of the interactions of VP40 with the WW domains of a human ubiquitin ligase, Nedd4. (D) Crystal structure of octameric VP40.

In 2003, I joined the Macromolecular Crystallography Group (now named the Structural Biology Group) at the European Synchrotron Radiation Facility (ESRF) on a European collaborative project (FP6 project: EURAMAN) to study the GlcNAc-phosphotransferase

enzyme, responsible for targeting of proteins to lysosomes and associated with lysosomal storage diseases. The project focused on the gamma-subunit of this enzyme, believed to be required for specific recognition of lysosomal proteins. Initial work was carried out in Norway (one month visit) and the subsequent development of a large-scale overexpression system in mammalian cells (293T cells) was achieved in Grenoble. Successful expression and purification of this protein allowed us to raise specific antibodies, which could then be used for further biochemical assays (search for binding partners) and cellular localisation studies (electron microscopy using gold-labelled antibodies). The project is still ongoing and is being carried out by collaborators in Norway and Finland.

During my post-doctoral work at ESRF, I rapidly became involved in the ongoing Structural genomics project that focused on a radiation-resistant bacterium, *Deinococcus radiodurans*. This Gram-positive, red-pigmented, non-sporulating and non-pathogenic bacterium (Figure 2) can withstand very high doses of ionising radiation (over 10,000 Gy), which introduce several hundred double-strand breaks into its genome (Cox and Battista 2005).

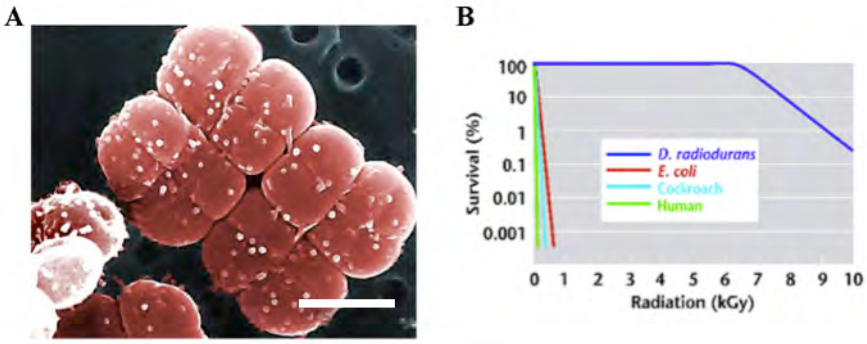


Figure 2: (A) Scanning Electron Microscope image of Deinococcus radiodurans (false-colour, by Peggy A. O'Cain and Margaret C. Henk, Louisiana State; modified by Peter Reid, The University of Edinburgh). The scale bar represents 1µM. (B) Typical survival curves for D. radiodurans, E. coli, human and cockroach.

The robustness of this bacterium is due to a combination of factors: (i) cell structure and nucleoid organisation (Levin-Zaidman, Englander et al. 2003), (ii) strong oxidative stress resistance mechanisms (including high Manganese content) (Daly, Gaidamakova et al. 2007; Slade and Radman 2011) and (iii) an efficient and precise DNA repair system (Zahradka, Slade et al. 2006). This bacterium was of particular interest to a team working in a synchrotron – a powerful source of ionising radiation, in this case X-rays.

The main goal of the project in 2003 was to determine the crystal structures of selected protein targets (approximately 100 proteins) from *D. radiodurans* (see section B.1. below), but other methods available at the ESRF were also used and/or further developed to characterise this unusual bacterium (see section A.1. and A.2.). As the years passed, my involvement in the project increased and by 2006 I was the principal coordinator of the *D. radiodurans* research project and my interest focused on its remarkably efficient DNA repair repertoire (see section C).

A. *D. radiodurans* and synchrotron-based experiments

1. Irradiation experiments

In 2006, we made use of the ID17 medical beamline at ESRF to carry out an X-ray irradiation experiment on *D. radiodurans* cells. *D. radiodurans* is a mesophile and is conventionally grown at 30ºC in rich medium with aeration. For this experiment, the cells were grown in either rich M53 medium or in minimal medium (MM) supplemented with Manganese and were subjected to various doses of X-rays: 0, 0.1, 0.5, 1, 5, 10, 15, 20 and 30 kGy (Figure 3). After irradiation, the cells were grown again at 30ºC and samples were collected at 0, 3, 6 and 24 hours post-irradiation for further analysis.

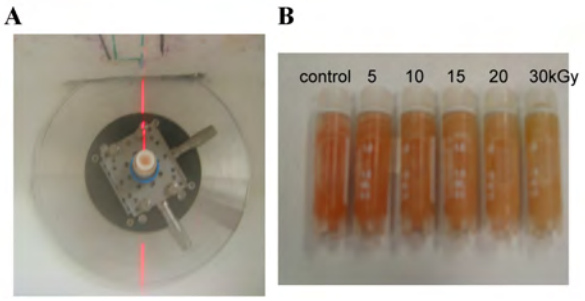


Figure 3: (A) Sample environment on ID17 beamline at ESRF. (B) 2 ml samples of D. radiodurans cells following irradiation with 0, 5, 10, 15, 20 and 30 kGy. The pink colour of D. radiodurans cells is due to the presence of deinoxanthin (a carotenoid). Irradiation of the cells significantly alters the colour of the cells.

The goal of this experiment was to investigate the survival curve of *D. radiodurans* in different growth environments as a function of X-ray dose (Figure 4) and to produce cell samples for analysis of DNA damage and protein expression levels as a function of dose and recovery time post-irradiation.

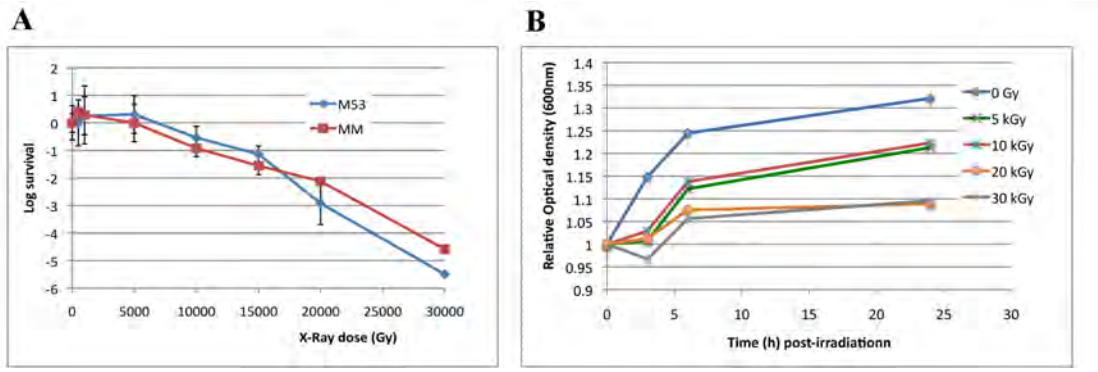


Figure 4: (A) Survival curves of D. radiodurans cells grown in rich M53 medium or minimal medium (MM) as a function of X-ray dose. The plotted data points correspond to the average of three independent measurements, except in the case of the highest dose (30 kGy) where only a single measurement could be made. (B) Growth curves of the non-irradiated and irradiated cells. The optical density at 600nm was measured at 0, 3, 6 and 24 hours after the irradiation of the cells.

As observed previously, our experiments confirmed that *D. radiodurans* withstands extremely high doses of ionising radiation and in this particular case, X-rays. Up to 5 kGy the viability of the irradiated cells is comparable to those of control cells and it is only above 10 kGy that the viability starts to be affected. Our growth curves of the control and irradiated

cells also show that the cells arrest their growth for the first 3 hours post-irradiation, time during which they repair their largely fragmented genomes. Between 3 and 6 hours, upon completion of DNA repair, cells start growing again at rates similar to control cells.

2. Cell imaging

D. radiodurans is usually found in the form of diads or tetrads with an average cell diameter of 1 µm. Among the many advanced biological imaging methods available to date, there is currently no method that can image intact a-few-micron-thick samples in three dimensions (3D), reaching the resolution of 10 nanometres or better, without the risk of structural artefacts. The recently developed X-ray diffraction microscopy (XDM) technique has the potential to reach this goal (Miao, Charalambous et al. 1999). The high penetration power of X-rays allows probing thick samples without sectioning, while lens limitations, such as low efficiency or depth of focus, are no longer limiting factors to achieve high-resolution 3D imaging. XDM is an imaging method based on diffraction. When the diffracted wave field from a sample is known at the far field, the image reconstruction is straightforward by an inverse Fourier transformation. Experimentally, only the amplitude of the wave field is available, as the phase is not measured. Phase retrieval algorithms have been developed that impose a priori known information as constraints iteratively to find a solution (Elser 2003). However, the convergence of the phasing algorithm is strongly dependent on the quality of the diffraction pattern. The general procedure outlined above has been used very successfully with radiation-hard samples, but obtaining quality data from weakly scattering biological samples, especially in the frozen state, has been challenging (Sayre 2008). For successful imaging, one needs to preserve the samples in the amorphous ice state throughout sample handling and data collection.

In collaboration with scientists from the ID10C beamline at ESRF, we used *D. radiodurans* cells to develop cryo-XDM, which combined plunge-freezing of samples with data collection in the cryogenic-temperature gas environment (Figure 5A), as routinely used on macromolecular crystallography beamlines. For these experiments, *D. radiodurans* cells were grown in TGY medium at 30ºC for 3 to 4 hours and exponentially growing cells were resuspended in 10% glycerol just before plunge freezing in liquid ethane. The cells were frozen in nylon sample loops. The ID10C beamline was set-up to produce coherent 8 keV X-ray illumination on the sample and the diffraction data were collected on a charge-coupled device (CCD).

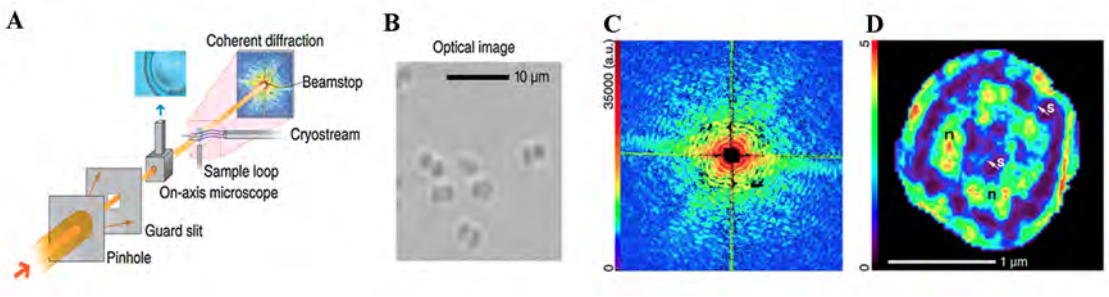


Figure 5: (A) Experimental setup at ID10C, ESRF. The shaded region indicates a vacuum environment. The non-vacuum sample environment is a 40 cm long section of the beam path housing slits, a sample goniometer, an on-axis visualization video microscope by MAATEL, and an Oxford Cryostream cooler 700 series. (B) Optical microscope image of D. radiodurans with 320x magnification. (C) An assembled diffraction pattern of a frozen-hydrated D. radiodurans tetrad. The total exposure time is 7 min using focused 8 keV X-rays. The measured diffraction signals at the edge of the array extend to 30 nm spatial half period

in the sample. (D) 2D reconstruction of D. radiodurans. The arrows at ''s'' highlight a diagonal structure which may be the septum and the areas labelled with ''n'' indicate possible nucleoid regions.

Figure 5C shows a full diffraction pattern measured from frozen-hydrated, unstained *D. radiodurans*. The estimated radiation dose is ~$3x10^7$ Gy and no signs of radiation damage were observed. The reconstruction shown in Figure 5D was obtained by averaging the final reconstructions from 10 different random starts. The sample size is about 1.5 μm, calculated from the speckle size and reconstruction. Close observation suggests that the optically denser areas (labelled as ''n'') might be the nucleoid regions of the bacteria while ''s'' points out what may be the septa dividing the whole cell. The resolution in the current reconstruction is estimated to be 30 to 50 nm where features down to 30 nm are visible with a good contrast, but we believe that the technique can be refined to achieve higher resolution in the near future. This work was published in Physical Review Letters in 2009 (Lima, Wiegart et al. 2009).

B. Structural genomics approach

1. High-throughput methods

At the end of the 1990's almost simultaneous explosions in the techniques of protein production and purification, synchrotron-based macromolecular crystallography and bioinformatics led to the proposal of a new field of structural biology called structural genomics (SG) (Terwilliger 2000). At the core of SG projects was the perceived potential to rapidly produce, purify, crystallise and solve the crystal structures of many thousands of gene products from many different organisms. Significant investments in software developments and rapid access to synchrotron radiation facilities meant that structure elucidation was already much more straightforward than previously and so the major focus of SG became the ability to control high yield protein production procedures to provide material for X-ray crystallography.

The structural genomics project initiated at the ESRF in 2002 aimed to solve structures of protein targets from *D. radiodurans* associated with its radiation resistance. Target selection was based on the available literature and knowledge of biological responses to irradiation, oxidative damage and desiccation. These criteria led to an initial set of proteins with *a priori* relevance either as part of a known functional category or as a protein unique to *D. radiodurans*. Membrane proteins were excluded from this study. Additionally, some gene products potentially allowing derivative-free crystal structure determination were selected. In these cases gene products were chosen based on their high sulphur content (at least 4% sulphur-containing residues) and a set of ten were chosen based on the putative presence of metals in the native protein (Figure 6).

Initial cloning, expression and solubility assessment of these ~100 targets were outsourced to a start-up biotechnology company in Grenoble, Protein'eXpert (now named PX' Therapeutics). All targets were cloned into a bacterial expression vector that produced proteins fused to a non-cleavable N-terminal His-tag. Expression was tested in different *E. coli* strains at either 37ºC or 15ºC in a highly automated manner. With this automatic screening process approximately 70% of the targets cloned produced visible expression (as tested on SDS-PAGE) at 37°C using *E. coli* BL21(DE3) and 50% produced soluble proteins. Whenever successful, these clones were then used by the ESRF Structural Biology (SB)

Group in order to pursue large-scale expression to produce sufficient quantities of protein for crystallisation trials.

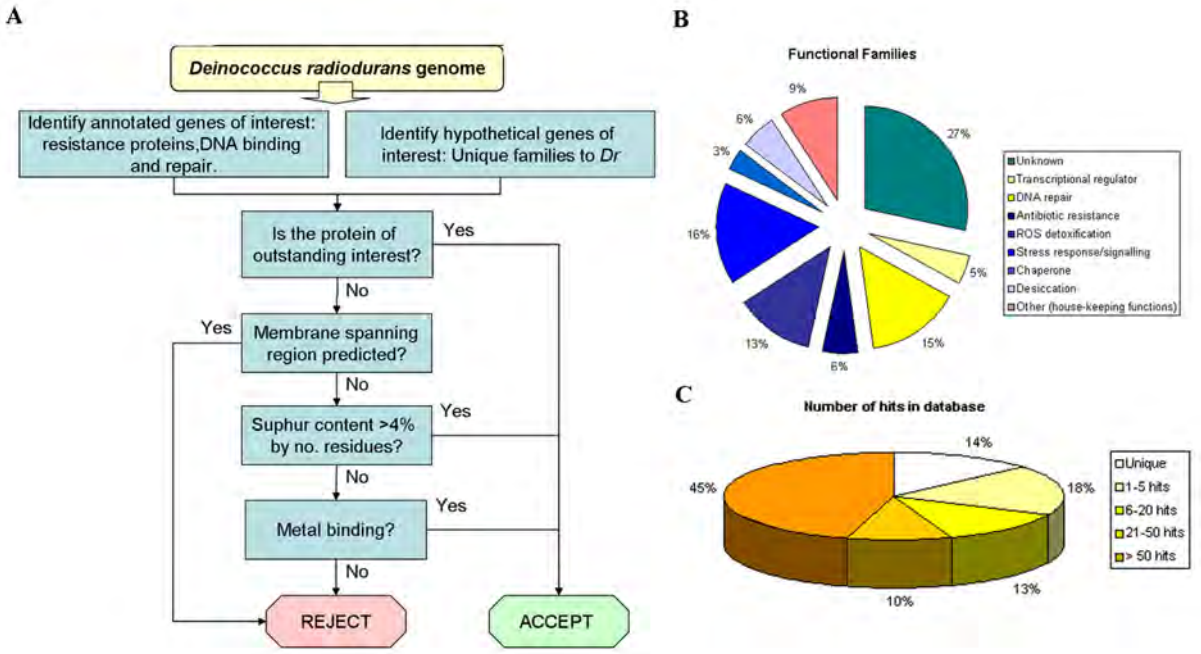


Figure 6: (A) Flowchart illustrating the process used to select ~100 targets from the D. radiodurans genome. (B) Functional distribution of these target proteins. (C) Frequency with which these targets were found in other bacterial genomes.

In 2004, a round of manual expression and solubility tests were carried out at the ESRF in order to validate the results of the automatic screening. This rapidly revealed significant differences in results, showing a generally higher level of expression and solubility compared to those obtained using the automated micro batch technique. We therefore decided to carry out systematic small-scale (manual) expression and solubility testing of all the clones (24 in total), which had not produced soluble proteins in the automatic screening. These tests were conducted at 20°C and various bacterial strains were evaluated. Both the levels of expression and solubility increased significantly following this manual screening procedure and these tests revealed that soluble protein could be obtained in at least one cell line for all of these recalcitrant targets.

We then subjected all these targets to medium-scale protein production (0.5 L cultures). In these, soluble product was obtained in more than 70% of the cases tested. This is rather less than the 100% success rate predicted in the small-scale tests and the differences observed in the production of soluble protein between the small- and medium-scale cultures are most likely the result of two phenomena: (i) an over-estimation of the soluble protein in the small-scale expression tests due to the use of SDS-PAGE as analysis tool and (ii) the significant changes in bacterial growth conditions when increasing sample volume from 5 ml to a 0.5 L culture, both of which are known to affect the level of expression and solubility of recombinant proteins (Sugar, Jenney et al. 2005). In over one third of the clones tested, a sufficient amount of protein was obtained from this medium-scale culture to carry out additional purification steps and set up crystallisation trials.

In 90% of cases, at least two purification steps were required to obtain the homogeneity suitable for crystallisation trials. A number of quality control steps were used to

ensure the protein sample was pure and homogeneous. Analytical size-exclusion chromatography was routinely used (in over 75% of cases) to assess both the quality of the protein samples and to investigate the oligomeric states of the targets. The dual role of gel filtration as a purification step and as a quality control/biochemical characterisation step makes it a very attractive tool when handling several proteins in parallel. Dynamic light scattering was also routinely used to check the polydispersity of the samples prior to crystallisation trials.

Crystallisation has been the most serious bottleneck encountered in SG projects. In this study, 33 proteins were expressed and purified in sufficient quantities for crystallisation to be attempted. A coarse screen of crystallisation conditions was set up using commercially available screens and initial hits optimised by a finer screen around the initial condition. For more than half the proteins studied robotic screening was used for initial crystallisation trials. However, in some cases manual screening resulted in crystallisation leads where robotic screening failed. This observation may be due to differences in drop handling and equilibration kinetics during the crystallisation process. The robot sets up 0.2 μl drops using the sitting-drop method, while the drops set up manually were larger (2 μl) and used the hanging-drop method. Of the 33 proteins that reached the crystallisation stage, 13 (40%) produced diffraction quality crystals (Figure 7). This ratio is similar to that observed in other laboratory scale structural genomics initiatives (Terwilliger 2000; Quevillon-Cheruel, Dominique et al. 2004; Segelke, Schafer et al. 2004; Moreland, Ashton et al. 2005) and can be explained by the trial-and-error, rather than strictly rational, approach to crystallisation used in this and other similar studies. A more systematic examination of the crystallisation phase diagram of each of the targets would possibly have increased our yield of crystallised proteins.

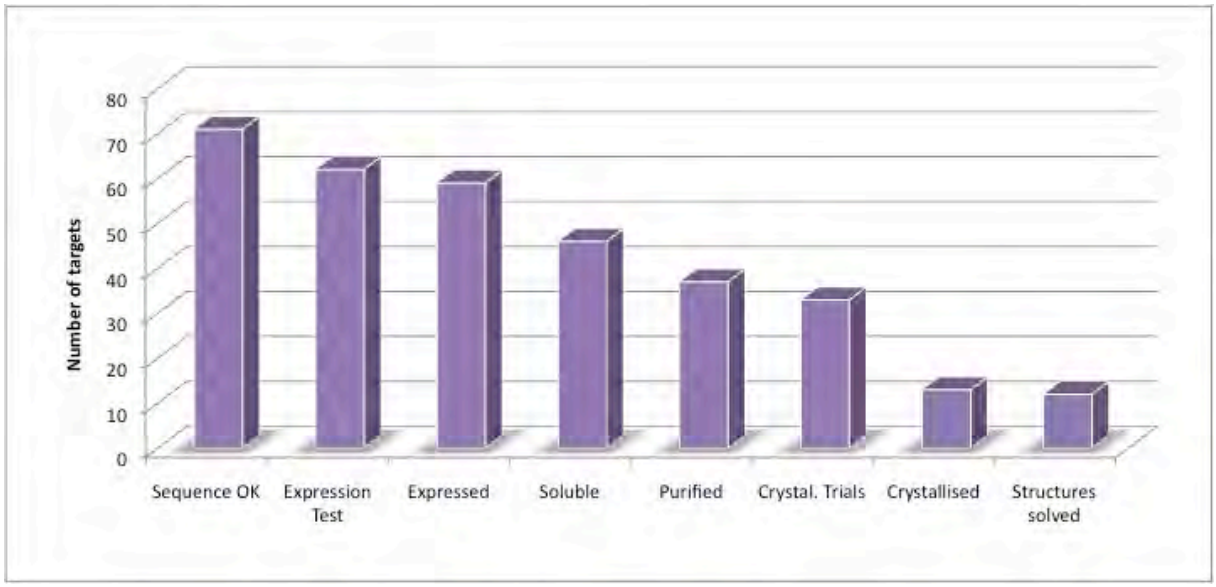


Figure 7: Status of the laboratory-scale D. radiodurans structural genomics project in 2006. The major bottleneck was the crystallisation step.

The structure determination process was clearly not a limiting factor in this work and undoubtedly the proximity to synchrotron beamlines was essential in achieving this. Of the 13 proteins crystallised, 12 have resulted in successful structure elucidation (Figure 7). The main structure determination method employed was by far experimental phasing using single-wavelength anomalous dispersion (SAD). Five structures were determined by exploiting

metal ions intrinsic to the native protein (Zn^{2+} or Fe^{2+}), three were solved using selenomethionine protein derivatives and one structure determination exploited sulphur as the anomalous scatterer. One of the 13 structures was solved using SIRAS, where a mercury compound was soaked into crystals and the three remaining structures were phased using molecular replacement methods.

By 2006, the project had successfully cloned 71 and expressed 59 genes resulting in the determination of 13 protein structures (Leiros, Kozielski-Stuhrmann et al. 2004; Meunier-Jamin, Kapp et al. 2004; Leiros, Moe et al. 2005; Leiros, Timmins et al. 2005; Timmins, Leiros et al. 2005; Dennis, Micossi et al. 2006; Moe, Leiros et al. 2006; Romao, Mitchell et al. 2006; Cuypers, Mitchell et al. 2007; Leiros and McSweeney 2007; Kapp, Macedo et al. 2008), which compares well with other, larger scale projects. During the evolution of this project the emphasis of the study shifted from a strictly structural genomics project into a scientific evaluation of a number of important and grouped genes. This had the consequence that low hanging fruit were sacrificed in order to allow projects of greater scientific interest to be pursued. The technology available was re-deployed to allow more intensive investigation in a project-orientated approach.

2. Crystal structure of sugar-bound MTHase

Within the context of the *D. radiodurans* structural genomics project, in 2003 I crystallised and determined the structure of an enzyme involved in trehalose synthesis, named maltooligosyltrehalose trehalohydrolase (or MTHase). This work was published in Journal of Molecular Biology in 2005 (Timmins, Leiros et al. 2005).

Trehalose is a non-reducing α-1,1-linked disaccharide (α-D-glucopyranosyl-1,1-α-D-glucopyranose) that accumulates in organisms under stress conditions. Trehalose appears to be the most effective stabiliser of proteins and membranes during prolonged dehydration by serving as a water substitute (Elbein, Pan et al. 2003). Studies have also revealed that trehalose can act as a free radical scavenger and thus protect proteins and DNA from damage caused by oxygen radicals. Since there is no terrestrial environment that generates high doses of radiation, the radiation resistance of *D. radiodurans* has been proposed to have arisen by selection for desiccation resistance (Mattimore and Battista 1996). The removal of water from a cell is a severe, often lethal stress. During desiccation, cell metabolism is arrested completely and subsequently needs to be fully recovered during rehydration, and extensive damage is caused to DNA (including double-strand breaks), proteins and lipid membranes. Damage to DNA and proteins is mediated mostly through reactive oxygen species.

At least three different pathways for the biological synthesis of trehalose have been described. The best-studied pathway involving the enzyme trehalose phosphate synthase is missing in *D. radiodurans*. Instead, *D. radiodurans* appears to use two alternative pathways for trehalose biosynthesis. It possesses the gene encoding trehalose synthase (DR2036), which has been shown to catalyse the conversion of the α-1,4-glycosidic linkage of maltose to the α-1,1-glycosidic bond of trehalose. In addition, *D. radiodurans* has two extra genes for trehalose metabolism, which encode a maltooligosyltrehalose synthase (DR0463; MTSase) and a maltooligosyltrehalose trehalohydrolase (DR0464; MTHase). Together these proteins catalyse a two-step reaction that breaks down maltooligosaccharides or starch into trehalose (Figure 8A). MTSase converts the glycosidic bond between the last two glucose moieties from an α-1,4-linkage to an α-1,1-linkage, thus producing a non-reducing glycosyl trehaloside. The second enzyme (MTHase) cleaves the α-1,4-linkage adjacent to the α-1,1-glycosidic bond to release free trehalose.

D. radiodurans MTHase was expressed in BL21 (DE3) Star cells with an N-terminal His-tag and was purified on a Ni-chelating sepharose, followed by a size-exclusion chromatography. Large crystals (600 µm x 150 µm x 20 µm) were obtained using the hanging-drop method in 18-15% (w/v) PEG 2000 MME, 0.1 M Tris-HCl pH 8.5, 0.05-0.2 M $MgCl_2$ and 10% glycerol. A high resolution (1.1 Å) data set was collected on ID29 at ESRF (Table 1). The crystals belonged to the orthorhombic space group $P2_12_12_1$ with cell dimensions a=59.6Å, b=66.6Å and c=152.5Å and contained one molecule per asymmetric unit. For structure determination, seleno-methionine-substituted protein was prepared that crystallised in the same conditions as the native and these crystals were used for a SAD experiment performed at the selenium-absorption edge on ID29. Anomalous data to a resolution of 2.0Å were collected (Table 1). Crystals containing either MTHase-trehalose or MTHase-maltose complexes, were obtained in similar conditions as the apo-enzyme, but with either 25% (w/v) trehalose or 20% (w/v) maltose added to the mother liquor and data was collected on these crystals to respectively 1.5 and 1.2 Å (Table 1).

Table 1*: Statistics from the data collections.*

X-ray statistics	**SeMet-MTHase**	**Native MTHase**	**MTHase-trehalose**	**MTHase-maltose**
Beamline	ID29	ID29	ID14-EH2	ID14-EH2
Wavelength (Å)	0.979	1.005	0.933	0.933
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
PDB Entry Code	-	2BHU	2BHY	2BHZ
Unit cell (Å)	a=59.82 b=66.51 c=152.01	a=59.58 b=66.62 c=152.51	a=60.03 b=66.56 c=153.14	a=59.14 b=66.56 c=153.50
Resolution (Å)	55.90-2.00 (2.11-2.00)	46.93-1.10 (1.16-1.10)	20.00-1.50 (1.58-1.50)	20.00-1.20 (1.26-1.20)
No. of unique reflections	41 475	216 600	98 045	187 653
Multiplicity	3.6 (3.3)	2.6 (1.8)	3.4 (2.7)	3.8 (2.7)
Completeness (%)	99.2 (95.8)	88.3 (82.9)	99.1 (95.8)	99.2 (95.2)
Mean ($\langle I \rangle / \langle \sigma_I \rangle$)	21.2 (15.9)	7.1 (2.3)	14.8 (3.9)	12.6 (2.5)
R_{sym} (%) [a]	4.6 (5.6)	8.2 (38.6)	4.8 (26.4)	5.9 (39.3)
R_{anom} (%) [b]	6.1 (7.5)			
FOM_{SAD} [c]	0.554			
FOM_{SF} [d]	0.84 (1.6Å)			

The numbers in parentheses represent values in the highest of 10 resolution shells, and the resolution limits for these are also indicated.

[a] $R_{sym} = (\sum_h \sum_i | I_i(h) - \langle I(h) \rangle |) / (\sum_h \sum_i I_i(h))$, where $I_i(h)$ is the *i*th measurement of reflection *h* and $\langle I(h) \rangle$ is the weighted mean of all measurements of *h*.

[b] $R_{anom} = \sum(|\langle I+ \rangle - \langle I- \rangle|) / (\sum(\langle I+ \rangle + \langle I- \rangle))$, where $\langle I \rangle$ is the mean intensity of the reflection.

[c] FOM_{SAD} = Figure of merit after SAD phasing

[d] FOM_{SF} = Figure of merit after solvent flattening and phase extension to 1.6 Å using an additional native data set

MTHase is a monomer and displays three major domains (Figure 8B): an N-terminal domain (domain N) that consists of an eight-stranded immunoglobulin-like Greek key fold , a

central catalytic $(\beta/\alpha)_8$ barrel domain (domain A) and a C-terminal domain (domain C) that consists of a six-stranded γ-crystallin-type fold. The N domain is connected to domain A via a long, fully extended linker and packs tightly against the N-terminal side of the $(\beta/\alpha)_8$ barrel, while domain C packs against the C-terminal side of the $(\beta/\alpha)_8$ barrel. Three additional subdomains (domain B, D and E) protrude from the central domain A (Figure 9). The domains are named according to α-amylase nomenclature (MacGregor, Janecek et al. 2001). Subdomain E, which is missing in most members of the α-amylase family, consists of a long α-helix (Eα1) in MTHase, followed by two smaller ones (Eα2 and Eα3) that connect Aβ8 and Aα8 of domain A. This long α-helix (Eα1) lies above the active site and the loop connecting this helix to the following shorter helices is missing in the final refined model, implying that this structural unit is highly flexible. Three strictly conserved carboxylic acid residues are found in family 13 glycosidases, which have been shown by mutagenesis studies to be the essential catalytic residues (Sogaard, Kadziola et al. 1993). Sequence and structural comparison have revealed that these residues correspond to Asp275, Glu308 and Asp400 in MTHase. These catalytic residues are situated on the top of the $(\beta/\alpha)_8$ barrel. Asp275 and Glu308 are located at the C-terminal tips of β-strands Aβ4 and Aβ5 respectively, while Asp400 is contributed by a short α-helix linking Aβ7 to Aα7. The positions of the waters in the active site pocket of MTHase are highly conserved.

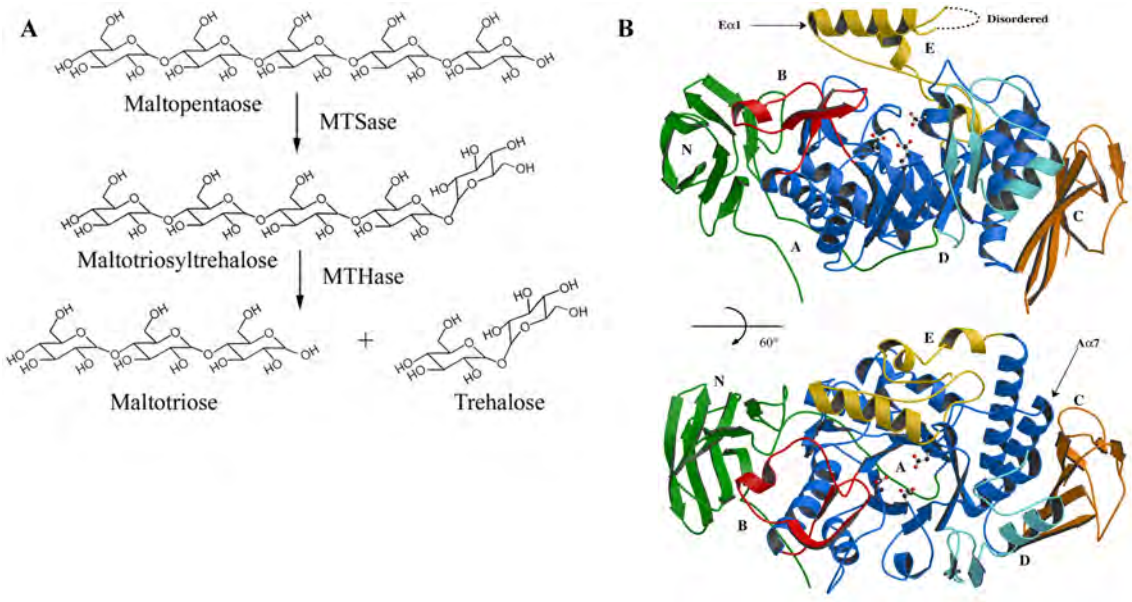


Figure 8: (A) Schematic diagram of the two-step breakdown of maltopentaose into trehalose and maltotriose catalysed by MTSase and MTHase. (B) Ribbon representations of the apo-form of MTHase. Each domain is represented in a different colour and is labelled. The catalytic residues are illustrated in ball-and-sticks. The disordered loop (residues 473-481) is shown as a dashed line and Aα7 and Eα1 helices are indicated by arrows.

Comparison of our crystal structures obtained for the apo- and sugar-bound MTHase reveals that the overall conformation of MTHase in complex with sugars is essentially the same as the native structure. The MTHase-trehalose complex structure was solved to 1.5Å resolution and seven trehalose molecules and two single glucose rings could be built into the electron density maps. Four of the trehalose molecules are involved in crystal packing. One trehalose molecule (Tre1) sits at the entry of the channel leading to the active site and a second trehalose molecule and two single glucose rings could be modelled into electron density at the exit of the substrate binding cleft, thus indicating a possible pathway for product

release. Most importantly, a trehalose molecule (Tre2) could be modelled into the +1 and +2 subsites of the substrate binding cleft (Figure 9B). The major difference between a maltooligosaccharide and a maltooligotrehaloside substrate is the orientation of the terminal glucose moiety, implying that the +2 subsite is the most critical region for specific substrate recognition (Figure 9A). This is indeed the case in the MTHase-trehalose complex. The presence of trehalose (Tre2) in the +1 and +2 subsites of the MTHase-trehalose complex has allowed the identification of the residues involved in the selective binding of the α-1,1-linked sugar, trehalose, as opposed to other maltooligosaccharides. These crucial residues are located in the A and D domains. The sugar units in the +1 and +2 subsites are both specifically recognised by an extensive hydrogen-bonding network involving many water molecules. A majority of the residues involved in the specific recognition of the sugar ring in the +2 site are conserved within the MTHase family. Importantly, the His332 Nε2 atom hydrogen-bonds directly to the O6 atoms of the two sugar units of the bound trehalose molecule. This interaction is trehalose specific, since the presence of both O6 atoms at hydrogen-bonding distance from His332 Nε2 could only occur with a α-1,1-linked disaccharide.

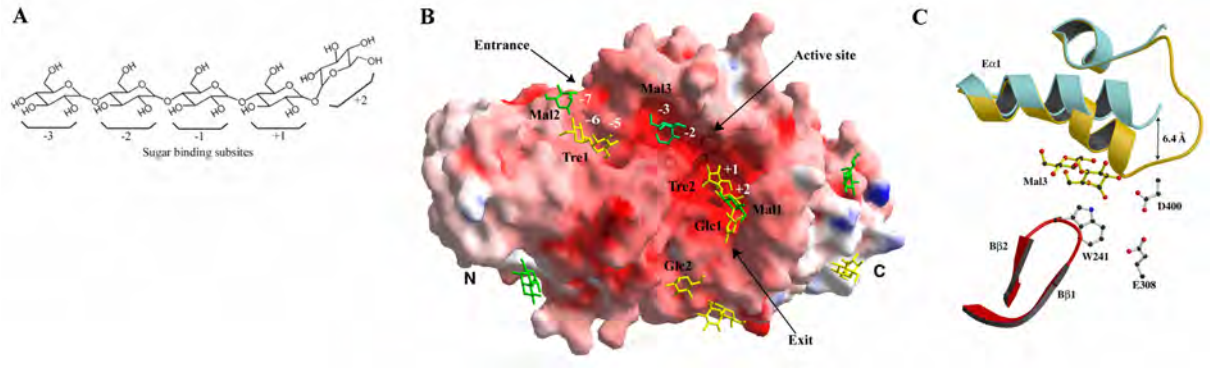


Figure 9: (A) Schematic diagram of the maltooligotrehaloside-binding subsites in MTHase. (B) Illustration of the electrostatic surface potential of MTHase in its substrate-bound form (complexed with maltose), contoured at ±20 kT/e in which red and blue represent negative and positive potentials respectively. The bound sugars are presented in ball-and-stick models in yellow (trehalose) and green (maltose), and arrows point to the entry and exit channels. The sugar-binding subsites are labelled in white. The secondary structure elements, Eα1, Bβ1 and Bβ2, are visible through the electrostatic potential surface. (C) Ribbon representation of the conformational change in subdomain E of MTHase upon binding to Mal3 (native conformation in blue and maltose-bound conformation in yellow).

The structure of MTHase in complex with maltose (an α-1,4-linked disaccharide, which mimics a short portion of a maltooligosaccharide chain) was solved to 1.2Å resolution. As in the case of trehalose, the final refined model contains seven sugar molecules. Three of these maltose molecules are involved in crystal packing, two of which are found in similar positions to the trehalose. Another maltose molecule makes extensive contacts with the long linker that joins the N domain to the A domain and with the bottom of the $(\beta/\alpha)_8$ barrel. This interaction is most likely contributing to the stability of the protein, but may not be biologically relevant, since the binding site is far from the active site. Additionally, there is a maltose (Mal1) in the +2 and +3 subsites (Figure 9B). The ring in the +2 site makes similar hydrogen-bonding interactions as the trehalose ring at the equivalent position, however the maltose ring is not stabilised by the hydrophobic stacking against His310. The +3 subsite is located at the exit of the active site, in close proximity to where the glucose moiety was found in the MTHase-trehalose complex. Moreover, two maltose moieties were found in the entry

channel leading to the active site. One of these molecules (Mal2) is located very close to the entry site, with one of its rings closely overlapping with the trehalose molecule at the equivalent position (corresponding to the -6 subsite). The second maltose (Mal3), however, is tightly bound at the -2 and -3 subsites. For this interaction to take place, the long α-helix of subdomain E (Eα1) undergoes a significant conformational change in which it closes down on the maltose, in order to trap it in the channel leading to the active site. This new conformation of Eα1 is now kinked at residue 464, resulting in a 20° bend in the helix and a 6.4 Å-displacement of the Cα atom of Phe472 (Figure 9C).

The three-dimensional crystal structures of MTHase in its apo-form and in complex with both maltose and trehalose revealed the mode of substrate recognition of MTHase. This is most likely shared by other members of the subfamily of MTHase enzymes, since most of the critical residues are conserved. Unlike other members of the glycosidase family 13, MTHase undergoes a significant conformational change upon substrate binding, in order to guide the sugar chain into the active site. The presence of this unusual trehalose synthesis pathway in the extreme radiation and desiccation-resistant bacterium *D. radiodurans* suggests that this additional pathway may be essential, especially as the 'classical' pathway involving trehalose-phosphate synthase is missing. The rapid breakdown of the most widely available source of maltodextrins in nature, *i.e.* soluble starch, would be an efficient mechanism to rapidly produce trehalose in response to environmental stress.

Crystals of seleno-methionine-substituted MTHase were later used for a study of the effect of X-ray dose rate on radiation damage during macromolecular crystallography data collection. This work led to a publication in 2006 (Leiros, Timmins et al. 2006).

C. DNA Repair in *D. radiodurans*

1. Double-strand break repair

1.1. RecO protein

In 2004, within the context of the ongoing *D. radiodurans* structural genomics project, I began to work on the expression, purification and crystallisation of RecO, an essential DNA repair protein involved in the repair of the most lethal type of DNA damage, namely double-strand breaks. In prokaryotes, homologous recombination, in addition to its fundamental role in genetic diversification of bacterial genomes, plays an essential role in the repair of a variety of DNA damage, including double-strand breaks (Figure 10) (Kuzminov 1999; Michel, Grompone et al. 2004). In *E. coli*, the initiation of homologous recombination can be carried out by either the RecBCD or the RecFOR proteins; in both cases these proteins act as mediators for RecA binding to single-stranded DNA (ssDNA) in order to allow for homologous strand invasion (Kowalczykowski, Dixon et al. 1994). Comparative studies of bacterial genomes have revealed that many genomes display an incomplete set of DNA repair systems. While RecBCD has been shown to be the major DNA recombination pathway in *E. coli*, the RecFOR pathway actually appears to be the more frequent pathway in bacterial genomes (Rocha, Cornet et al. 2005). RecR, along with RecA and the resolvases, have been found to be nearly ubiquitous in bacteria, suggesting they must be playing essential roles. RecO and RecF are less well conserved and appear to be missing in a number of species.

In contrast to the (*E. coli*) RecBCD pathway the mechanism of recombinational repair mediated by the RecFOR pathway is still only poorly understood. The major difference between these two recombination pathways lies in the initiation step, also known as

presynapsis. RecFOR displaces ssDNA binding protein (SSB) prior to loading of RecA onto the ssDNA, while RecBCD directly catalyses the latter step. In addition, RecBCD is known to function as a single holoenzyme (Singleton, Dillingham et al. 2004), whereas it is still unclear whether RecFOR exists as a functional complex *in vivo* and how it mediates presynapsis. It has been shown that RecF, RecO and RecR are all required for protecting the nascent lagging strand when replication forks are stalled on UV-radiation induced damage sites. The absence of any of the RecFOR genes causes *E. coli* to become both hypersensitive to UV radiation and to display extensive degradation in the nascent lagging strand, suggesting that they form an epistatic group (Chow and Courcelle 2004).

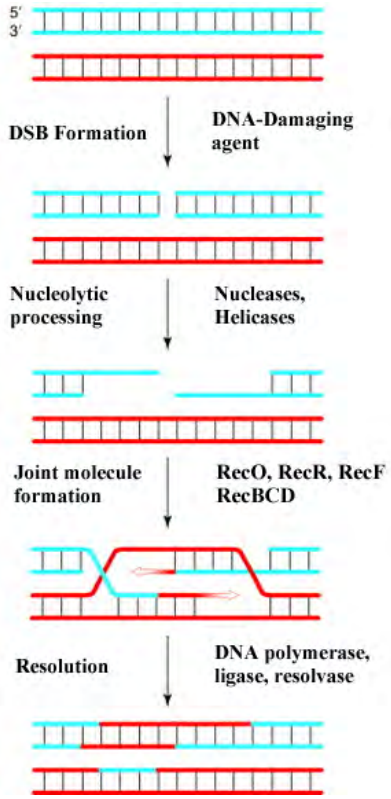


Figure 10: DNA double-strand break (DSB) repair through homologous recombination. DSBs introduced by DNA damaging agents are processed to a single-stranded region with a 3′ overhang by a nuclease, a helicase or both. RecA is loaded by RecBCD or RecFOR onto the single-stranded DNA to form a nucleoprotein filament that searches for the homologous duplex DNA. After the search has been successfully completed, DNA strand exchange generates a joint molecule between the homologous damaged and undamaged duplex DNAs. DNA synthesis, requiring a DNA polymerase, its accessory factors and a ligase, restores the missing information. Finally, resolution of crossed DNA strands (Holliday junctions) by a resolvase yields two intact duplex DNAs. (Figure adapted from (Kanaar, Hoeijmakers et al. 1998)).

RecR from *D. radiodurans* has been found to form tetramers in a ring-like structure (Lee, Kim et al. 2004). A central hole in the tetramer has a diameter of 30 Å, and is suited for binding double-stranded DNA (dsDNA). In the crystal structure of RecR, two such tetramers are interlocked and it is therefore believed that RecR tetramers can open and close in order to bind DNA. The RecFOR complex is specific for the dsDNA-ssDNA junction in the lagging strand of DNA at a stalled replication fork. Mapping this specificity into the individual components of the RecFOR complex is crucial in understanding the ongoing processes at such sites. As a first step in understanding the structural aspects of its involvement in DNA

repair, we determined the first crystal structure of a RecO protein (Figure 11) and identified the regions of RecO involved in DNA binding (Figure 12). This work was published in EMBO Journal in 2005 (Leiros, Timmins et al. 2005).

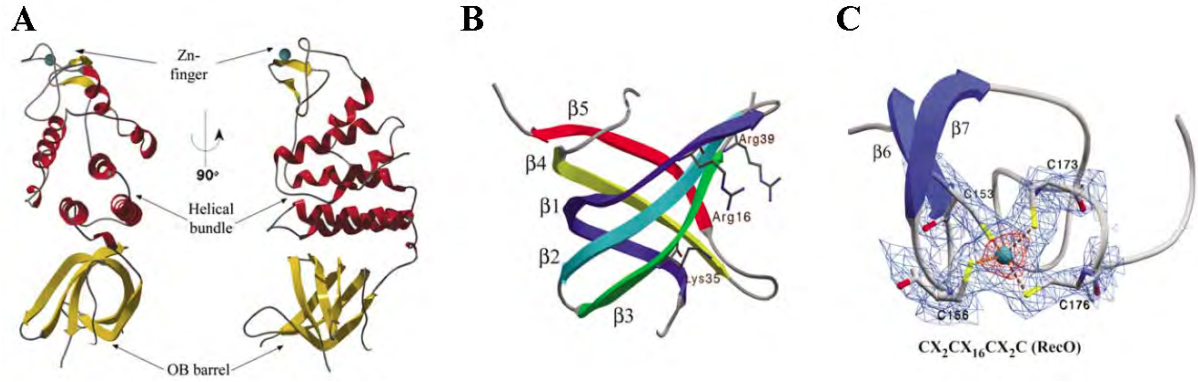


Figure 11: (A) Ribbon illustrations of the secondary structure elements in RecO. α-Helices are shown in red and β-strands in yellow. The zinc atom in the zinc-finger motif is illustrated as a cyan sphere. (B) OB-fold of RecO, coloured in rainbow colours ranging from blue at the N-terminus to red at the C-terminus. The nucleotide-interacting residues of OB domains are generally found in the cleft on the right-hand side of the displayed domain. Residues in the β1–β2 loop and in the β4–β5 loop line this cleft. (C) Ribbon illustration of the zinc-finger in RecO. The four cysteine residues coordinating the zinc atom are illustrated in a ball-and-stick presentation and the zinc atom is shown as a cyan sphere. The electron density maps are 2mFo-DFC maps contoured at 2σ (blue) and 7σ (red), respectively. The cysteine residues and the zinc atom were all omitted from the refinement.

D. radiodurans RecO (*dr*RecO) was expressed in *E. coli* with an N-terminal His-tag, purified and crystallised in either 0.2 M calcium acetate, 8% (w/v) PEG 20,000, 8% (w/v) PEG 550 MME, buffered with 0.1 M Tris-HCl at pH 8.5, or 0.8 M sodium formate, 10% (w/v) PEG 8000, 10% (w/v) PEG 1000, buffered with 0.1 M Tris-HCl at pH 7.5. A first data set was collected to 2.4 Å resolution and indicated the presence of an anomalous scatterer in the crystal. A XANES (X-ray absorption near edge structure) analysis on ID14-4 at the ESRF confirmed this scatterer to be zinc. The structure was subsequently solved by the SAD method using data collected to 2.7 Å at the zinc absorption edge (Table 2). *dr*RecO is composed of an N-terminal oligonucleotide/oligosaccharide binding (OB) fold region, a three-helix bundle, a Cys_4 zinc finger motif and finally a group of four helices spatially inserted between the three-helix bundle and the zinc finger motif. The OB fold domain and zinc finger motif reside at opposite ends of the protein with the α-helices clustered in the central part (Figure 11A).

Using only the OB fold domain as a search template, DALI gave good scores for several nucleotide-binding proteins, with the best match being for human and *E. coli* SSB proteins, followed by the breast cancer susceptibility gene 2 encoded protein (BRCA2), which also aligns well to the OB fold domain of *dr*RecO. Within the OB fold, ssDNA binding is generally mediated by three distinct structural elements: the β1–β2 loop, the β4–β5 loop and the C-terminal part of β3, where the first provides positively charged residues to form ion-pair(s) with the ssDNA phosphodiester backbone and the latter two each provide one conserved aromatic moiety (Bochkarev and Bochkareva 2004) (Figure 11B). In *dr*RecO, positively charged residues occupy similar positions. In all the compared SSB proteins, there are at least two aromatic residues that have been shown to be important in ssDNA binding. In *dr*RecO, no such aromatic residues can be found and there could be a tendency for

substitution of aromatic residues in SSB into positively charged residues in *dr*RecO. Among RecO proteins, R16 and K35 in *dr*RecO are conserved in being positively charged, thus further strengthening this possibility.

Table 2: *Data collection, structure solution and refinement statistics. Numbers in parenthesis are for data with a high-resolution cut-off at 4Å. ♣: Total/Protein/Water/Zinc.*

Dataset	Peak	Remote
Data Collection statistics:		
Beamline	ID14-4	ID14-1
Wavelength (Å)	1.2827	0.934
Resolution range (Å)	19.9-2.7 (2.85-2.70)	48.5-2.4 (2.53-2.40)
R_{sym} (%)	9.3 (48.1)	6.2 (55.2)
Multiplicity	8.3 (8.3)	2.3 (2.3)
Mean I/σI	17.8 (4.1)	8.5 (1.7)
Completeness (%)	99.7 (100.0)	98.8 (99.7)
Space group	C2	C2
Unit cell parameters (a, b, c and β)	134.7Å, 52.1Å, 100.7Å and 106.4º	134.4Å, 52.4Å, 101.1Å and 106.3º
Refinement statistics:		
# Atoms ♣	-	3557/3532/23/2
B-factors ♣	-	53/56/43/46
R_{free} (%)	-	26.7 (36.0)
R_{work} (%)	-	22.5 (33.0)
Geometry:		
Bonds (Å)	-	0.010
Angles (°)	-	1.324

Zinc-fingers have long been considered as important mediators of protein–DNA interactions (Branden and Tooze 1991). Typically, they have the secondary structure succession β–β–α where the α-helix contacts DNA through major groove interactions. In *dr*RecO, the α-helix is not formed (Figure 11C) and the importance of this zinc-finger motif is questionable as the *E. coli* and *T. thermophilus* RecO proteins both lack the C-terminal zinc-finger motif found in many bacterial RecO proteins.

The DNA binding properties of *dr*RecO were subsequently investigated by electrophoretic mobility shift assays (EMSA) and an extensive mutational study of RecO (Figure 12). Non-sequence specific protein-DNA interactions are predominantly either of a hydrophobic character, where aromatic or aliphatic residues contact DNA bases, or of an ionic character, where polar or positively charged residues contact the negatively charged phosphate backbone. *E. coli* RecO has previously been shown to interact with both ssDNA and dsDNA (Luisi-DeLuca and Kolodner 1994). In our study, we show that *dr*RecO also interacts with both ssDNA and dsDNA oligonucleotides and with supercoiled plasmid DNA. Sequence alignments and secondary structure predictions of RecO proteins from other organisms indicated that the overall structure may be conserved even in RecO proteins that lack the zinc finger found in *dr*RecO, suggesting that the zinc finger motif in *dr*RecO may

serve mainly as a structural scaffold for maintaining the rigidity of the protein, rather than as a mediator of protein-protein or protein-DNA interactions. However, when EDTA was added to our EMSAs in order to disrupt the zinc finger, the DNA-binding was highly affected, although not totally abolished, indicating that the zinc-finger may contribute in part to *dr*RecO-DNA binding. Electrostatic surface potentials highlight two regions of *dr*RecO having pronounced positive charges, *i*) a ridge in the α5-α6 region, which is close to the zinc finger and non-conserved among RecO proteins, is exposed to solvent and very rich in positively charged residues, and *ii*) a region in the OB-barrel, which appears to be conserved between RecO proteins and is rich in positively charged residues (Figure 12C).

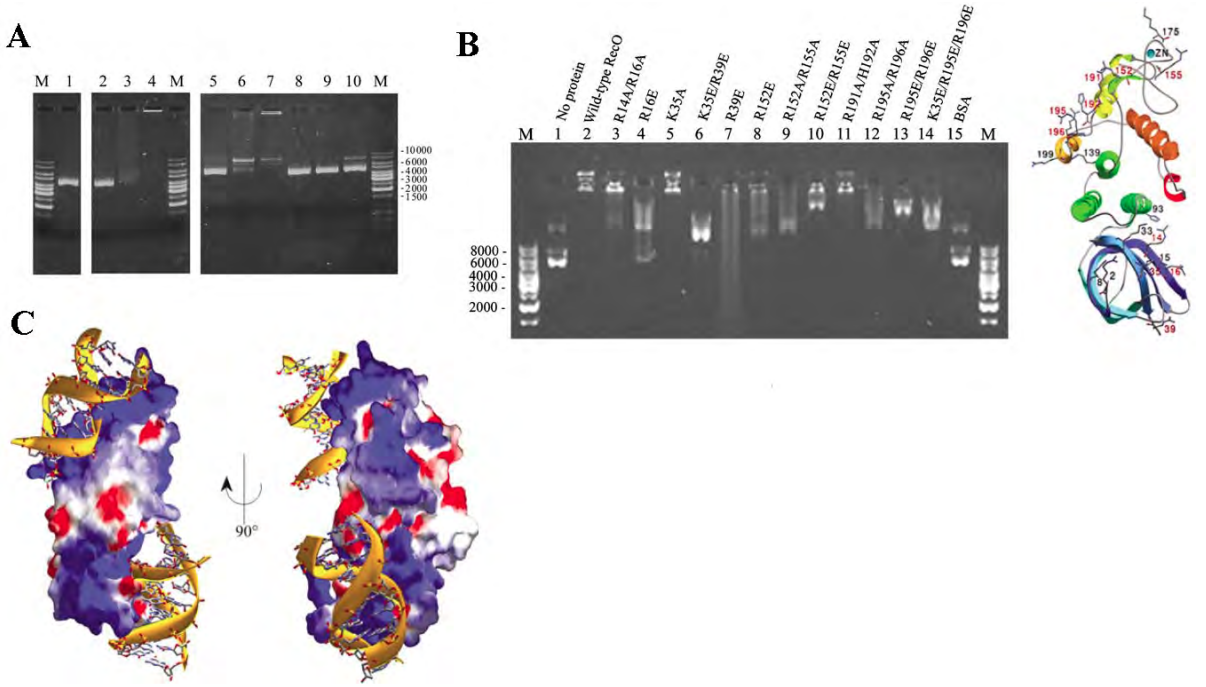


Figure 12: (A) EMSA of drRecO. Lane M, DNA size markers in base pairs; lane 1, 0.2 µg supercoiled plasmid DNA (pcDNA3.1); lane 2, pcDNA3.1 + 40mM $MgCl_2$; lane 3, pcDNA3.1 + 3 µg RecO; lane 4, pcDNA3.1 + RecO + 40m $MgCl_2$; lanes 5–7, as for lane 4 but using 0.05, 0.5 and 2 µg RecO, respectively; lane 8, as for lane 2; lane 9, pcDNA3.1 + 5 µg BSA; lane 10, pcDNA3.1 + 5 µg BSA + 40mM $MgCl_2$. (B) EMSA of mutant RecO proteins. Lane 1, 0.2 µg supercoiled plasmid DNA (pcDNA3.1); lane 2, pcDNA3.1 + 40mM $MgCl_2$ + 5 µg RecO; lanes 3–14, as for lane 2, but using 5 µg of the mutant indicated above the lanes; lane 15, as for lane 2, but using 5 µg BSA. The ribbon illustration on the right indicates the location of the mutations on the structure of drRecO. (C) Models for dsDNA interacting with RecO based on the DNA-binding studies and mutational analysis. Two alternative binding sites involving the OB barrel (bottom) and a positive patch (190-RHAVRRTVR-200) unique for drRecO ending at the zinc-finger (top) are shown.

Based on these findings, 16 mutants of *dr*RecO were made, of which 12 were successfully purified and used in DNA-binding studies. Positively charged residues in both of the indicated regions of the protein were mutated into Ala and/or Glu. As both positive regions cover a large area, mutating positively charged residues into Ala may not be sufficient to abolish all DNA-binding in these regions. Mutation into Glu was therefore chosen in some cases in order to generate repulsion between the mutant proteins and DNA, hopefully without compromising the stability of the protein. The results of the DNA-binding study on mutated *dr*RecO show that the DNA-binding affinity of mutant RecO proteins is reduced compared to wild-type *dr*RecO and in some cases, no residual DNA-binding was observed (Figure 12B).

The mutations with the most pronounced effect on DNA-binding were R16E, K35E/R39E and K35E/R195E/R196E, strongly indicating that the OB-fold region is particularly important for DNA binding in *dr*RecO, but that the second site is also involved to a lesser extent in DNA binding.

Since the RecFOR complex is specific for the ssDNA-dsDNA junction in a stalled replication fork, while the individual RecR and RecO proteins can interact with both ssDNA and dsDNA, formation of a molecular complex between the proteins appears to be necessary in order to obtain the specificity for the ssDNA-dsDNA junction in damaged DNA. In 2005, we therefore decided to focus our efforts on the study of the RecO-RecR protein complex from *D. radiodurans* and to gradually move away from the structural genomics approach.

1.2. RecOR complex

At the time, there was evidence that *E. coli* RecO and RecR formed a stable complex and together could remove SSB and allow RecA loading in the absence of RecF, suggesting that RecF is (at least partially) dispensable for this process (Umezu and Kolodner 1994). *E. coli* RecR is a dimer in solution (Hegde, Qin et al. 1996) and the two proteins are suggested to form a complex in a 1:1 molar ratio, presumably in the form of a 2:2 heterotetramer (Umezu and Kolodner 1994). In contrast, *dr*RecR has been shown to be a tetramer or an octamer in solution in a concentration dependent manner (Lee, Kim et al. 2004).

The gene encoding *D. radiodurans* RecR (*dr*RecR) had been cloned along with the *dr*RecO gene as part of the *D. radiodurans* structural genomics project into a bacterial expression vector with an N-terminal His-tag. To isolate the RecO-RecR (RecOR) complex, cells expressing *dr*RecO were mixed and lysed with cells expressing *dr*RecR, and the soluble fraction of the cell lysate was purified by Ni-affinity chromatography, followed by ion-exchange and size-exclusion chromatography (SEC). In the final SEC step, the RecOR complex eluted as a single peak (elution volume: 12.1 ml) corresponding to a molecular weight of approximately 150–160 kDa. RecO and RecR alone eluted later from the size exclusion column at 17.1 and 13.5 ml, respectively, in agreement with their calculated molecular weights (26 and 95 kDa). The ratio of *dr*RecR/*dr*RecO in the RecOR complex was estimated by comparison with known amounts of individually purified *dr*RecR and *dr*RecO on SDS–PAGE, and was found to be in a 2:1 ratio, probably in the form of a 4:2 heterohexamer, which best agrees with its molecular weight derived from SEC.

Table 3: *Data collection statistics.*

Data collection	
Space group	C2
Unit cell (Å/°)	a=144.0, b=83.2, c=66.6, β=106.9
Resolution range (Å)	45-3.8 (4.0-3.8)
Number of unique reflections	7194
Redundancy	3.1 (2.1)
R_{sym} (%)	13.6 (63.7)
Completeness (%)	95.9 (95.9)
Mean I/σ(I)	6.5 (1.4)
Wilson B-value ($Å^2$)	96.0

Numbers in parenthesis are for data with a high-resolution cut-off at 4Å

Crystals of the RecOR complex were obtained using the hanging-drop method by mixing 1 µl of protein at 12 mg/ml with 1 µl of reservoir solution (0.1 M MES pH 6.5, 10% dioxane, 1.6 M Ammonium sulphate). X-ray intensity data was collected on ID14-2 at the ESRF (Table 3). The crystal used for data collection belonged to the monoclinic space group C2 and had unit cell dimensions of a=144.0Å, b=83.2Å, c=66.6Å, β=106.9°. Diffraction was observed to a maximum resolution of about 3.8Å. Molecular replacement (MR) searches were performed using one monomer of *dr*RecR (pdb1VDD) and one monomer of *dr*RecO (pdb1W3S) as search models. Phaser found a unique solution with no symmetry-generated clashes. With a limited observation-to-parameter-ratio, refinement was terminated after a single round of rigid-body refinement (R_{work}/R_{free} of 45.9%/44.3%, respectively) and subsequent manual rebuilding. Although the *dr*RecOR complex structure is limited to 3.8 Å resolution, the corresponding maps are of overall good quality for the protein backbone, also revealing regions (mainly in RecR) that had to be remodelled due to complex formation.

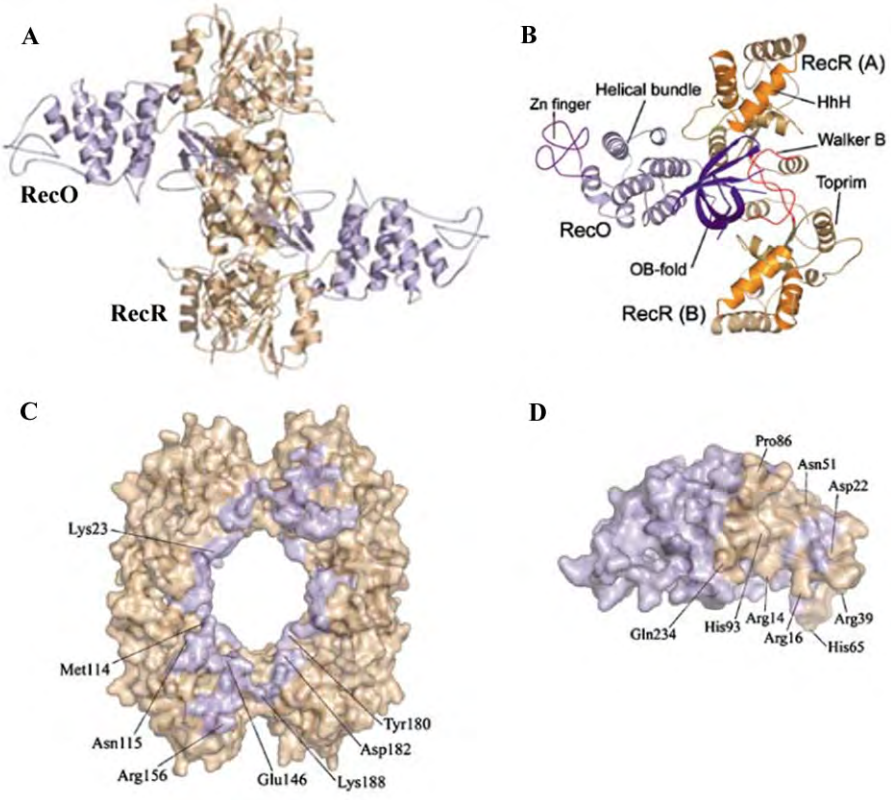


Figure 13: (A) Crystal structure of the hetero-hexameric drRecOR complex as reconsitituted by symmetry. drRecO is coloured in blue and drRecR molecules are coloured in gold. (B) Ribbon representation of the heterotrimeric drRecOR assembly present in the asymmetric unit. The structural elements of drRecO and drRecR are indicated. (C) Surface representation of drRecR indicating in blue the residues involved in protein-protein interactions with drRecO. Selected residues from the interface are labelled on one of the drRecR monomers. (D) Surface representation of drRecO indicating in gold the residues involved in protein-protein interactions with drRecR. Selected residues from the interface are labelled.

The *dr*RecOR complex has a 2:1 molecular ratio of *dr*RecR to *dr*RecO, where the content of the crystallographic asymmetric unit is a heterotrimer (Figure 13B). The C2 space group symmetry operator (-X,Y,-Z) generates the most probable biologically relevant heterohexameric molecular unit. By this symmetry operator, the tetrameric structure previously described for *dr*RecR (Lee, Kim et al. 2004) is reconstructed, and the estimated

molecular weight of such a complex is consistent with its SEC elution profile. This entity thus consists of a heterohexamer of two *dr*RecO molecules and four *dr*RecR molecules (Figure 13A). The *dr*RecO molecules are rather unexpectedly positioned on either side of the tetrameric ring of *dr*RecR, obstructing access to the interior of the ring. In this complex, the N-terminal OB-fold domains of each of the two *dr*RecO molecules point towards the interior of the *dr*RecR ring. The accessible surface area lost for one monomer of *dr*RecO when binding to the tetramer of *dr*RecR was calculated to be around 1500Å^2, or 14% of the total accessible surface area for a monomer of *dr*RecO, thus well within the representative range found for protein-protein complexes.

Most of the *dr*RecR residues contributing to complex formation are located around its central hole, while those of *dr*RecO are clustered on one side of the OB-fold domain and its neighbouring α-helix (Figure 13C and D). The OB-fold domain of *dr*RecO contacts residues from both the N-terminal Helix-hairpin-Helix (HhH) motif and the C-terminal region, including the Walker B motif of *drR*ecR. Both these regions in *dr*RecR participate in domain swapping with neighbouring molecules and are thus also critical for stabilisation of the tetrameric structure of *dr*RecR. *dr*RecO also contacts the Toprim domain of *dr*RecR. The interactions are mostly of hydrophobic character. Relatively few intermolecular hydrogen bonds or potential ion pairs were found. Only two of the latter were identified between *dr*RecO and the *dr*RecR tetramer: RecO-His93 to RecR-Glu146 (MolA) and RecO-His65 to RecR-Asp182 (MolD). All the regions of *dr*RecR responsible for interacting with *dr*RecO are strongly conserved in RecR proteins, including Glu146 and Asp182. In addition ionic interactions between residues Asp22 and Arg39 of symmetry-related *dr*RecO molecules also contribute to the overall stability of the biological complex, as they form *dr*RecO-*dr*RecO interactions across the interior of the ring.

To validate our low-resolution crystal structure and verify the importance of the ionic interaction between His93 from *dr*RecO and Glu146 from *dr*RecR, we mutated His93 to Glu. This H93E mutant *dr*RecO was no longer able to form a stable complex with wild-type *dr*RecR. *dr*RecO His93 and *dr*RecR Glu146 are therefore essential in maintaining the hetero-hexameric assembly. Glu146 is highly conserved in RecR proteins and contributes to a characteristic acidic surface patch found in Toprim domains (Aravind, Leipe et al. 1998). Recent work by Honda *et al* also identified the equivalent glutamate residue (Glu144) of *T. thermophilus* RecR (ttRecR) as being critical for its interaction with ttRecO (Honda, Inoue et al. 2006). In this study, mutation of ttRecR Glu144 to Ala resulted in a non-functional RecR protein that could no longer facilitate the loading of RecA onto SSB-coated ssDNA. Two independent studies thus show the importance of this ionic interaction for the stability of the RecOR complex, and consequently for its function in RecA loading onto ssDNA. While most of the contacts between *dr*RecR and *dr*RecO are of hydrophobic nature, the ionic interactions appear to contribute significantly to the overall stability of the complex. RecR, unlike RecO, is highly conserved throughout bacteria and the observation that the residues responsible for binding to RecO are also strongly conserved suggests that this protein-protein interface may well be observed in other bacteria.

Since both *dr*RecO and *dr*RecR had been shown to interact with plasmid DNA, we carried out similar EMSAs with the purified *dr*RecOR complex. Addition of *dr*RecOR to plasmid DNA led to a significant retardation of the plasmid DNA, which was further enhanced in the absence of Mg^{2+} (Figure 14A). Ten mutant complexes of *dr*RecOR were used for DNA binding studies. In the case of *dr*RecR, two potential DNA binding sites were mutated: (i) K23 and R27 located in the N-terminal HhH motif and (ii) D182 and E183 at the C-terminus of the Walker B motif. Single and double mutations of K23 and R27 to Glu

interfere with DNA binding, but only in the absence of Mg^{2+}. For all three of these complexes, however, DNA binding is maintained in the presence of Mg^{2+}. The double mutation of K23 and R27 to Ala, however, fully disrupts DNA binding even in the presence of Mg^{2+} (Figure 14B). This finding suggests that these two residues are essential for *dr*RecOR DNA binding, but that Mg^{2+} coordination by the introduced Glu residues is compensating, at least in part, for the K23E/R27E mutations. The second DNA binding site (D182/E183) was proposed by Lee *et al* to be involved in Mg^{2+} coordination, therefore mutations to both Arg and Ala were prepared. However, only the double mutant D182A/E183A was soluble and could be used for the DNA binding study. This mutant was found to retain DNA binding, leading to a significant band shift (Figure 14B) that is very similar to that obtained for wild-type *dr*RecOR complex.

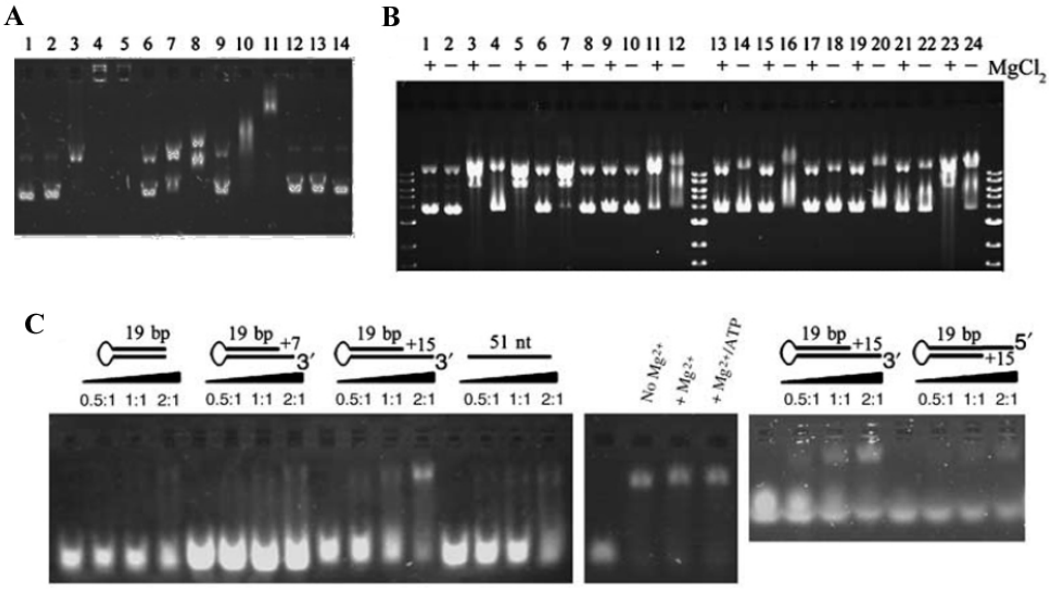


Figure 14: (A) EMSA analysis of drRecO, drRecR and drRecOR binding to supercoiled plasmid DNA. Lanes 1 and 2 correspond to plasmid DNA alone in the presence and absence of 40mM $MgCl_2$. 1 and 5 μg of each of the proteins were added to the DNA in the presence of 40mM $MgCl_2$ and separated by EMSA (lanes 3-4 for drRecO, lanes 6-7 for drRecR, lanes 9-10 for drRecOR and lanes 12-13 for BSA). In addition, in lanes 5, 8, 11 and 14, 5 μg of each of the proteins was mixed with plasmid in the absence of $MgCl_2$. (B) Mutant drRecOR-plasmid binding. Lanes 1 and 2 correspond to supercoiled plasmid DNA alone in the presence and absence of 40mM $MgCl_2$. Lanes 3-22 correspond to the analysis of the binding of drRecOR mutants (lanes 3, 4 wtRecO-RecR K23E; lanes 5, 6 wtRecO-RecR K27E; lanes 7, 8 wtRecO-RecR K23E/R27E; lanes 9, 10 wtRecO-RecR K23A/R27A; lanes 11, 12 wtRecO-RecR D182A/E183A; lanes 13, 14 RecO-R16E-wtRecR, lanes 15, 16 RecO-R39E-wtRecR, lanes 17, 18 RecO-R35E/R39E-wtRecR, lanes 19, 20 RecO-R35E/R195E/R196E-wtRecR, lanes 21, 22 RecO-R195E/R196E-wtRecR) to plasmid DNA, while lanes 23 and 24 correspond to the wild-type complex. All reactions were prepared both with (odd numbers) and without (even numbers) $MgCl_2$. (C) Four types of DNA (dsDNA, two 3'overhanging DNA, 3'-OH1 (+7nt) and 3'-OH2 (+15nt), and ssDNA) were mixed with wild-type drRecOR at various protein:DNA ratios (0.5:1, 1:1 and 2:1) and were analysed by EMSA. In the second panel drRecOR was mixed with 3'-OH2 at a 2:1 ratio in the absence and presence of either $MgCl_2$ alone or $MgCl_2$ and ATP. In the third panel drRecOR was mixed with 3'-OH2 and 5'-OH2 at various protein:DNA ratios (0.5:1, 1:1 and 2:1)

In addition, several complexes were prepared by combining mutant *dr*RecO proteins with wild-type *dr*RecR. These include mutations in the two previously proposed DNA binding regions of *dr*RecO: the OB-fold domain and the C-terminal positively charged ridge (Leiros, Timmins et al. 2005). Unlike the *dr*RecR mutants, DNA binding of these complexes is strongly affected by all the mutations in *dr*RecO, particularly in the presence of Mg^{2+} (Figure 14B). In particular, mutations in the OB-barrel completely abolish DNA binding in the presence of Mg^{2+}. Regardless of the presence or not of Mg^{2+}, the double OB-barrel (K35E/R39E) *dr*RecO mutant complex no longer binds DNA, while the *dr*RecO double mutant (R195E/R196E) seems to retain some DNA binding ability.

EMSA studies were also carried out with various DNA oligonucleotides, forming either single-stranded, double-stranded, 3'overhanging (OH1 and OH2 with respectively 7 base and 15 base overhangs) or 5'overhanging (15 base overhang) DNA substrates. *dr*RecO alone shows preferential binding to single-stranded DNA (and thus also binds to a lesser extent to both 3' and 5' overhanging DNA) and *dr*RecR shows no detectable binding to these short oligonucleotides (Figure 14C). Wild-type *dr*RecOR displays a clear preference for binding to 3'overhanging DNA compared to 5'overhanging, single- or double-stranded DNA. The highest affinity is seen for the longer 15-mer 3'overhang (Figure 14C). This binding was unaffected by the addition of the divalent cation, Mg^{2+}, and ATP. Four of the mutant *dr*RecOR complexes that showed severely impaired binding to plasmid DNA were tested for binding to single-stranded and double-stranded oligonucleotides and to the 15-mer 3'overhanging DNA substrate (OH2). Three of these four mutants here again displayed very weak binding to all DNA substrates. Unexpectedly, the mutant complex consisting of the double *dr*RecO mutant R195E/R196E associated with wild-type *dr*RecR exhibited an increased binding relative to wild-type *dr*RecOR (Figure 14C). None of the mutant complexes displayed a clear preference for one of the DNA substrates.

While the structure observed is consistent with the *dr*RecOR complex observed in solution, it is also unexpected. Lee *et al* proposed that *dr*RecR may act as a non-sliding DNA clamp, which could accommodate dsDNA within its central hole (Lee, Kim et al. 2004). In the *dr*RecOR complex, RecO reduces access to the interior of the ring, suggesting that such a complex, if functional, must most likely adopt a different mode of DNA binding. Based on our mutagenesis studies, within the RecOR complex both *dr*RecR and *dr*RecO provide key residues involved in binding to DNA; in *dr*RecR two positively charged amino acids, K23 and R27, located in the HhH motif are essential for *dr*RecOR-DNA interactions, as are two residues, K35 and R39, found in the OB-fold domain of *dr*RecO (Figure 15A). Interestingly, mutating one of these two sites was sufficient to fully disrupt plasmid DNA binding, suggesting that both sites are needed for stable association of *dr*RecOR with DNA. In both cases, these residues had been shown to be critical for the binding of the individual *dr*RecO and *dr*RecR proteins to dsDNA. In contrast additional regions of either *dr*RecR (such as Asp182) or *dr*RecO (positively-charged ridge at the C-terminus; residues Arg195 and Arg196) had been shown to be involved in DNA binding in the individual proteins, but only appear to be minor contributors to DNA binding within the context of the RecOR complex.

The preferred DNA substrate of the complex was identified as being dsDNA with a 3'overhang, such as those encountered at ssDNA-dsDNA junctions in the lagging strand of stalled replication forks. These junctions are believed to be the sites of action of RecF and RecOR *in vivo* (Morimatsu and Kowalczykowski 2003). The individual *dr*RecO and *dr*RecR proteins show a very different DNA binding pattern to that observed for *dr*RecOR. *dr*RecO alone displays preferential binding to ssDNA and *dr*RecR shows no detectable interaction with either ssDNA and dsDNA oligonucleotides under these conditions. In addition, mutating

the DNA binding sites of either *dr*RecO or *dr*RecR also resulted in mutant complexes that no longer displayed preferential binding to 3'overhanging DNA and retained weak binding to short fragments of both ssDNA and dsDNA. Selective binding to 3'overhanging DNA thus requires the assembly of RecO and RecR as a complex. This represents the first evidence that the *dr*RecOR complex by itself displays a clear preference for binding to 3'overhanging DNA in the absence of RecF. However, this does not exclude a role for RecF in stabilising the assembly on ssDNA-dsDNA junctions through its interaction with RecR *in vivo*.

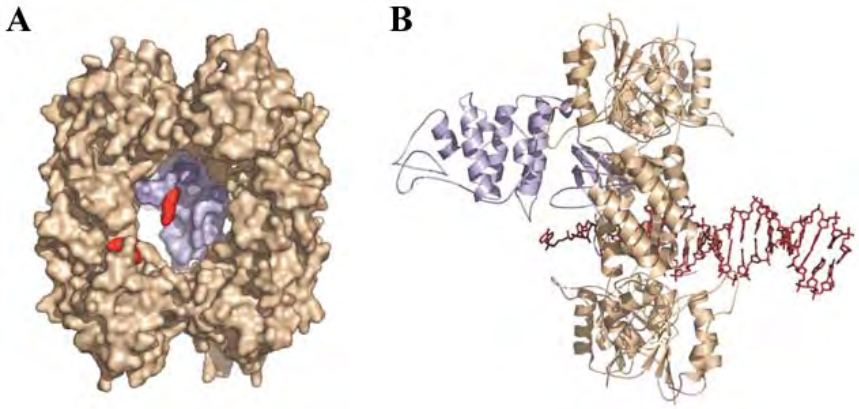


Figure 15: Models of the drRecOR complex bound to 3'-overhanging DNA. (A) Surface representation of the drRecOR complex after removal of one drRecO molecule (blue). Residues of drRecOR required for DNA binding are shown in red. (B) Ribbon representation of the drRecOR complex with one drRecO molecule replaced by a dsDNA fragment with a 3'overhanging strand.

These findings indicate that a structural rearrangement of the complex needs to take place in order for DNA to interact with residues located both inside the ring of *dr*RecR and in the OB-fold domain of *dr*RecO. We carried out chemical cross-linking and native gel electrophoresis studies of *dr*RecOR in the presence and absence of 3'overhanging dsDNA and these studies suggested that *dr*RecOR most likely undergoes both local conformational changes and a larger architectural reorganisation upon addition of DNA. Displacement of one of the two RecO molecules would leave enough space for a 3'overhanging strand of DNA to pass through the RecR ring and to interact with the OB-fold of the second RecO molecule (Figure 15B). In this model, the ssDNA-dsDNA junction interacts with the crucial residues from both RecO and RecR.

The steps involved in RecOR-mediated assembly of RecA onto ssDNA are not known in great detail. Upon assembly at ssDNA-dsDNA junctions, structural rearrangements may occur as a result of protein-protein and/or protein-DNA interactions between the various partners. The RecOR complex assembly may be affected by the interaction of RecR with RecF and RecO with SSB upon DNA binding. This may subsequently induce a conformational change in either RecR or RecO, resulting in the displacement of a RecO molecule and the stable assembly of RecOR on stalled replication forks. In the case of ttRecR it was recently shown that Glu144 plays an essential role in the formation of both the RecOR and the RecFR complexes (Honda, Inoue et al. 2006). RecO and RecF may thus be competing for a common binding site on RecR, in which case the relative affinity of RecO and RecF for RecR may be central to these events. This relative affinity may be species-specific since RecO and RecF proteins are less conserved. This process may also be regulated by the nucleotide-bound state of RecF in order to control the assembly of RecA onto ssDNA. The ATPase activity of RecF may provide the necessary energy required for such structural rearrangements. A better

understanding of the processes regulating complex formation between these partners will certainly help in establishing the detailed sequence of events leading to the assembly of RecA onto ssDNA. The three-dimensional structures of the individual proteins have served as solid frameworks to identify residues or regions potentially involved in either protein-protein or protein-DNA interactions. However, structural and biochemical studies of the protein complexes, such as that of *dr*RecOR, will now provide us with more valuable information regarding the functions of these proteins *in vivo*.

Table 4 *Data collection, refinement statistics*

Data collection		**Refinement**	
Space group	P1	Resolution (Å)	3.34
Cell dimensions		No. reflections	24006
a, *b*, *c* (Å)	63.1, 93.1, 92.3	R_{work} / R_{free} (%)	22.5 / 27.7
α, β, γ (°)	103.6.2, 110.4, 106.2	No. atoms	
Wavelength (Å)	0.9793	Protein	9156
Resolution (Å)	83.07 – 3.34 (3.52 – 3.34)	Zn	6
R_{sym} (%)	6.1 (35.1)	*B*-factors	
I / σI	8.8 (2.0)	Protein	106.5
Completeness (%)	94.2 (96.3)	Zn	96.7
Redundancy	1.8 (1.8)	R.m.s deviations	
		Bond lengths (Å)	0.009
		Bond angles (°)	1.259

*Values in parentheses are for highest-resolution shell.

This work was published in EMBO Journal in 2007 (Timmins, Leiros et al. 2007). In 2008, a post-doctoral fellow, Jens Radzimanowski, was recruited to pursue this work under my supervision. After many unsuccessful trials, in 2011 crystals of a mutant *dr*RecOR complex consisting of the double *dr*RecR mutant K23A/R27A associated with wild-type RecO were obtained that diffracted to a maximal resolution of 3.3 Å (Table 4). The structure was solved by molecular replacement using the single *dr*RecO and *dr*RecR structures as search models. The crystals belonged to space group P1 and contained one hetero-hexameric *dr*RecOR complex per asymmetric unit. This new crystal structure refined well using Phenix and the final geometry and statistics are very satisfactory (R_{fact}=22.5% and R_{free}=27.7%). In this crystal form *dr*RecOR reveals a more open conformation in comparison to the previous crystal structure (Figure 16A and B) and in this new conformation displacement of a RecO molecule is no longer a prerequisite for accommodation of ssDNA through the RecR ring. Interestingly, the important interaction between His93 from *dr*RecO and Glu146 from *dr*RecR is retained in this new crystal structure.

Extensive small-angle X-ray (SAXS) and neutron (SANS) scattering data were also collected on *dr*RecR and *dr*RecOR alone and in the presence of DNA in order to determine the solution structures of *dr*RecR and *dr*RecOR. These data revealed that *dr*RecR is actually a dimer in solution (Figure 16C and D) and not a tetramer or an octamer as previously reported (Lee, Kim et al. 2004). These higher order oligomers were most likely artifacts of the high protein concentrations used for crystallisation of RecR. The SAXS/SANS study of *dr*RecOR confirmed that *dr*RecOR was indeed a hetero-hexamer in solution and the solution structure agrees with the two available crystal structures, indicating that both of these conformations may occur in solution. These results indicate that tetramerisation of *dr*RecR may therefore be favoured by the binding to RecO. Finally, our SANS data suggest that the *dr*RecOR complex bound to ssDNA and dsDNA maintains its four-to-two stoichiometry and indicate that the

DNA is bound in the inside of the complex. A manuscript describing this multi-resolution study of the RecOR complex is in preparation (J. Radzimanowski *et al*, 2012).

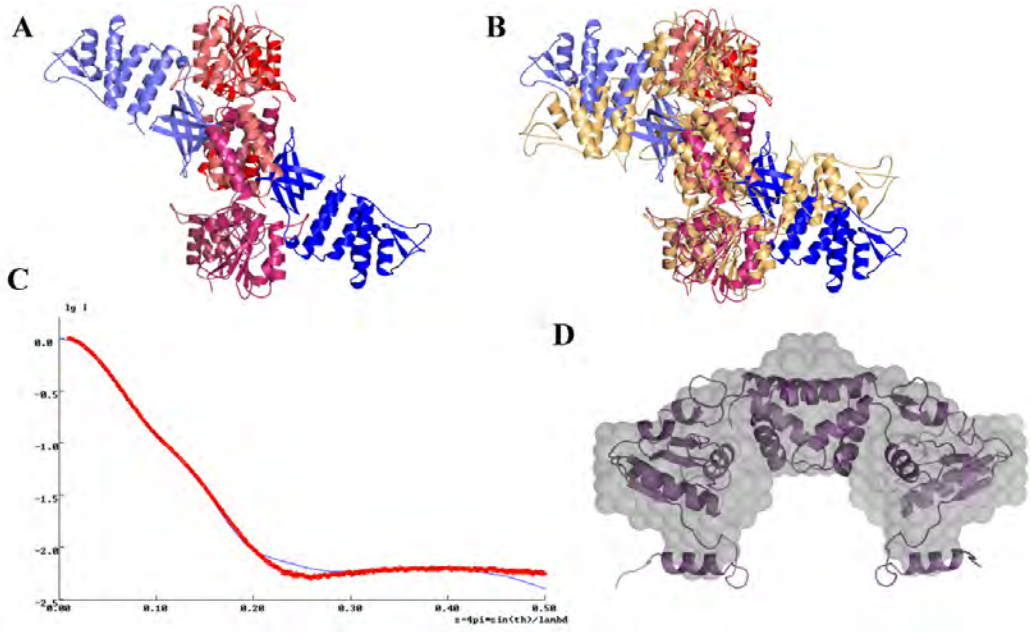


Figure 16: (A) Ribbon illustration of the structure of drRecOR (P1 space group). drRecR molecules are coloured pink and red, while drRecO proteins are coloured blue. (B) Overlay of the two drRecOR crystal structures. The P1 structure is coloured as in A, while the original C2 structure is coloured in gold. (C) Overlay of the experimental small-angle X-ray scattering curve obtained for drRecR (red) with the theoretical curve derived from a dimer of drRecR (blue line). (D) Ab initio model (space-filling model) derived from the SAXS data overlayed with the crystal structure of dimeric drRecR.

1.3. RecN protein

In 2008, Simone Pellegrino joined the ESRF Structural Biology Group as a PhD student to work under my supervision on the study of RecN, a protein proposed to be involved in the early steps of DSB repair. DSB recognition and the initial response to DNA damage are poorly characterised processes in bacteria. In eukaryotes, the central DSB response factor is the MRN complex consisting of Mre11/Rad50/Nbs1 (van den Bosch, Bree et al. 2003). This multifunctional complex triggers the cellular response to DNA damage, prepares the DNA ends for subsequent strand-exchange processes via its multiple nuclease activities and is implicated in tethering of DNA ends and chromatids (Lamarche, Orazio et al. 2010). Structural homologues of Mre11 and Rad50 are found in all three kingdoms of life but their roles do not appear to be conserved. Prokaryotic MR (Mre11/Rad50) complexes, also known as SbcCD complexes, have been implicated in processing of DNA ends and removal of DNA hairpin secondary structures during and after DNA replication (Cromie and Leach 2001; Connelly and Leach 2002). SbcC and its eukaryotic homologue, Rad50, belong to the SMC (structural maintenance of chromosomes) family that includes proteins involved in sister chromosome cohesion, chromosome condensation and chromosome segregation (Graumann 2001; Hirano 2005). SMC proteins share a common domain organisation in which the N and C-terminal domains associate to form a globular ATP binding cassette (ABC)-like nucleotide binding domain (NBD) and the central region forms a long coiled-coil (Figure 17A). Different organisms possess different numbers of SMC-like proteins. In addition to SbcC, many bacteria possess two additional SMC-like proteins, MukB and RecN (Graumann

and Knust 2009). Although the *recN* gene was originally identified in *E. coli*, it is found in most bacterial genomes. In *E. coli*, *recN* mutants display increased sensitivity to UV and γ-irradiation and to the DNA damaging agent Mitomycin C (Sargentini and Smith 1986; Lloyd, Porton et al. 1988). There is now increasing evidence that RecN plays an essential role in DSB repair in several bacterial species (Funayama, Narumi et al. 1999; Kidane, Sanchez et al. 2004; Kosa, Zdraveski et al. 2004; Wang and Maier 2008).

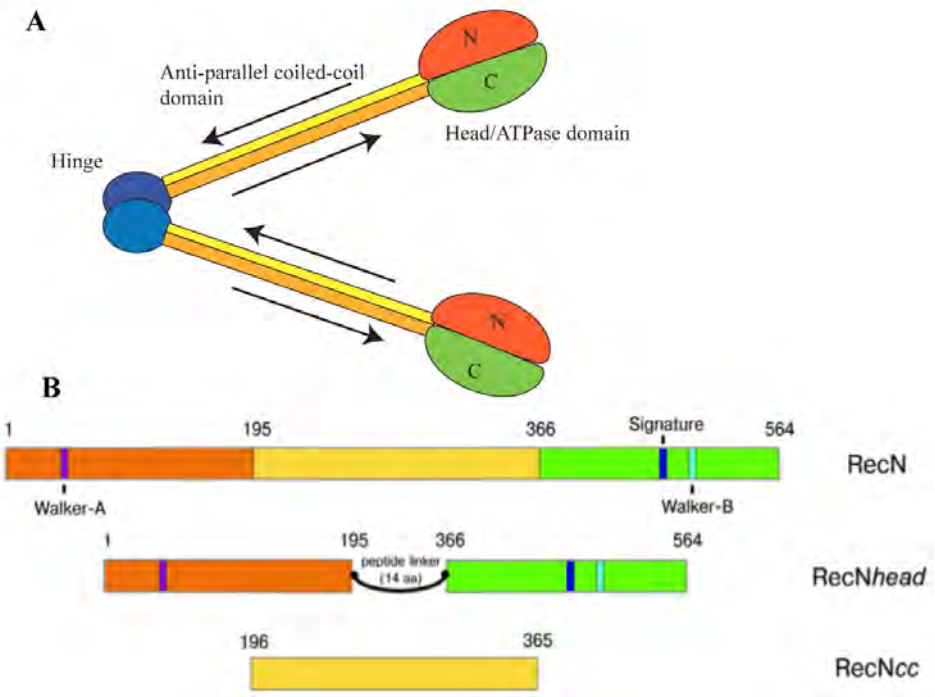


Figure 17: (A) Schematic representation of SMC proteins. They consist of a globular Head domain corresponding to the N- and C-terminal domains, which exhibits ATPase activity, an anti-parallel coiled-coil region and hinge domain involved in SMC dimerisation. (B) Schematic diagram of the constructs used for structural, biophysical and biochemical analysis of RecN. The conserved ATP binidng motifs (Walker A, Walker B and signature sequence) are indicated. The N-terminal domain is coloured orange, the C-terminal domain green and the coiled-coil region yellow.

However, the molecular mechanisms underlying these processes are still only poorly understood and at the start of S. Pellegrino's project, there was no structural information available for RecN. Bacterial RecN proteins show an overall conservation in length and domain organisation, and display a clear set of conserved functional motifs, typical of ABC proteins: Walker A, Walker B and signature sequence (Figure 17B). RecN has indeed been shown to display weak ATPase activity that is stimulated in most cases by the addition of DNA (Sanchez and Alonso 2005; Grove, Wood et al. 2009; Reyes, Patidar et al. 2010). *Bacillus subtilis* RecN can bind to discrete ssDNA ends and increase the local concentration of such ends, leading to the formation of rosette-like structures (Sanchez, Cardenas et al. 2008). To provide a structural framework for DSB recognition in bacteria, we carried out a high and low-resolution structural study of *D. radiodurans* RecN (*dr*RecN) and characterised its *in vitro* activity. A manuscript describing this work is in revision in Structure (Pellegrino *et al*, 2012) and S. Pellegrino successfully defended his PhD thesis in February 2012.

RecN was cloned into a bacterial expression vector for expression with a cleavable N-terminal His-tag. RecN was expressed in *E. coli* and purified by Ni-NTA chromatography

followed by SEC in a buffer containing high salt (1M NaCl), a requirement for the stability of the protein. After many unsuccessful crystallisation trials, we decided to design alternative constructs of RecN to determine the structures of its individual domains. Bioinformatics and limited proteolysis experiments followed by mass spectrometry analysis allowed us to define the putative boundaries of RecN's domains. Of these constructs, the fused N- and C-terminal domains with a 14-amino acid linker, named RecN*Head* and the coiled-coil construct containing residues 196-365, named RecN*cc* were successfully expressed and purified, yielding protein quantities and quality suitable for crystallisation trials (Figure 17B). The expression, purification, crystallisation and SAXS analysis of RecN*Head* and RecN*cc* were reported in two articles published in Acta Crystallographica Section F in 2012 (Pellegrino, de Sanctis et al. 2012; Pellegrino, Radzimanowski et al. 2012).

Table 5 *Data collection, phasing and refinement statistics.* *Values in parentheses are for highest-resolution shell.

	RecN*head* SeMet	**RecN*cc* Native**	**RecN*cc* SeMet**	**RecN*Δdd***
Data collection				
Space group	P 2_1	P 2_1	P 2_1	I 2
Cell dimensions				
a, *b*, *c* (Å)	129.9, 62.0, 133.8	73.2, 44.0, 133.6	71.6, 45.4, 133.4	72.2, 62.5, 145.5
α, β, γ (°)	90.0, 102.7, 90.0	90.0, 97.7, 90.0	90.0, 97.5, 90.0	90.0, 98.4, 90.0
Wavelength (Å)	0.9795 *(peak)*	0.992	0.9795 *(peak)*	0.873
Resolution (Å)	50.00 – 3.00 (3.16 – 3.00)	52.5 – 2.04 (2.15 – 2.04)	66.2 – 2.28 (2.40 – 2.28)	47.03-4.00 (4.15-4.00)
R_{sym} (%)	12.4 (41.9)	7.0 (32.1)	7.4 (22.8)	28.4 (75.8)
$I / \sigma I$	9.2 (3.3)	8.4 (2.8)	12.4 (5.2)	3.8 (1.5)
Completeness (%)	99.6 (100)	97.1 (98.2)	99.6 (99.7)	97.5 (93.8)
Redundancy	5.1 (5.2)	2.4 (2.3)	4.4 (4.4)	3.0 (2.9)
Refinement				
Resolution (Å)	49.5-3.00	45.9-2.04		47.0-4.00
No. reflections	39868	49848		5182
R_{work} / R_{free}† (%)	20.9/ 24.9	20.3 / 25.1		32.5 / 34.7
No. atoms				
Protein	10333	5019		3416
Water	116	444		0
B-factors				
Protein	43.49	31.74		70
Water	34.78	36.48		-
R.m.s deviations				
Bond lengths (Å)	0.004	0.003		0.003
Bond angles (°)	0.785	0.663		0.694

Crystals of RecN*Head* were obtained by mixing 1µl of protein at 10mg/ml with 1µl of reservoir solution (0.1M Tris-HCl pH 7.5 and 25% PEG 3,350). Diffraction data were collected on ID23-2 at ESRF to a maximum resolution of 3.0 Å. The structure was solved using SAD data collected at the Se absorption edge on ID14-4 using crystals of

selenomethionine substituted protein diffracting also to a maximum resolution of 3.0 Å. The structure was built and refined using Coot and Phenix (Table 5). The asymmetric unit contains four RecN*Head* molecules. The RecN*Head* construct was designed in order to reconstitute the complete nucleotide-binding domain (NBD) of RecN. Structural analysis revealed that this construct also included some additional regions (Figure 18): the coiled-coil anchor motif, the beginning of the predicted coiled-coil region and a short C-terminal helix-turn-helix (HTH) motif, which is unusual for Rad50 and SMC proteins. The overall structural organisation of RecN*Head* is similar to those of the head domains of *S. cerevisiae* SMC, *T. maritima* SMC and *P. furiosus* Rad50 proteins. RecN*Head* is a monomer both in solution (as seen by SAXS and multi-angle laser light scattering) and in the crystals.

The NBD of RecN*Head* consists of two lobes, in which lobe I, formed primarily by the N-terminal domain of RecN, adopts a typical Rossmann fold, consisting of a six-stranded anti-parallel β-sheet (strands β1-β2 and β4-β7) wrapping around the Walker A containing α-helix 1. Lobe II consists mostly of the C-terminal region of RecN and folds as a β–α–β sandwich in which helices α10-α12 pack against a mixed β-sheet (β3, β8, β12-β15). The N and C-terminal domains are held together by hydrogen-bonding interactions between β3 and β8 strands from the N-terminal domain and β12-β14 strands located in the C-terminal domain. The Walker-B motif is located on strand β12 and helix α10 contains the signature sequence, important for driving ATP-dependent NBD dimerisation (Altenberg 2003). The NBDs of RecN and other SMC or SMC-like proteins for which crystal structures are available only share a low overall sequence conservation (14% identity between RecN and pfRad50), but the residues responsible for ATP-binding (K67 in the Walker-A motif) and hydrolysis (D471 and E472 in the Walker-B motif) are strictly conserved and are located at similar positions in order to create half of the functional nucleotide binding pocket (Hopfner, Karcher et al. 2000). The signature sequence is also highly conserved, but unlike the Walker-A and B motifs, adopts a very different orientation, as a result of the rotation of the coiled-coil anchor motif relative to the NBD (Figure 18).

The anchor motif is primarily composed of a three-stranded anti-parallel β-sheet (β9-β11), helix α4 and two flexible linkers connecting β8 to α4, through a short helical turn (α3), and β11 to α11. Comparison with pfRad50 head domain reveals that the region connecting α2 to α4 including β8, which is substituted by a short helix in Rad50, is poorly conserved. As a result the entire coiled-coil anchor motif, along with the signature sequence of RecN*Head*, undergo a ~60° rotation relative to their position in pfRad50. Helix α4 plays an essential role in guiding the positioning of the helices belonging to the coiled-coil region (α5 and α10) so as to be roughly orthogonal to the NBD.

In order to investigate the ATP binding and hydrolysis activity of RecN*Head*, several point mutants were prepared targeting either the Walker A, Walker B or both motifs. A double mutant of RecN*Head* (RecN*Head*$^{K67A/E472Q}$) turned out to be of particular interest. Its elution profile from the SEC column, supported by dynamic light scattering (DLS) and multi-angle laser light scattering (MALLS) measurements, indicated that it was dimeric both in the presence and absence of nucleotide. It also displayed a strong absorbance at 260nm (with a 260/280 nm ratio of approximately 1), suggesting the presence of a bound nucleotide. SAXS measurements were performed on RecN*Head*$^{K67A/E472Q}$ and the scattering curve (Figure 18) obtained from these measurements could be used to successfully build an *ab initio* model of the overall envelope (Figure 18). The Guinier approximation estimated the radius of gyration (Rg) to be 3.45 nm and the molecular weight (~70 kDa) derived from the scattering at zero angle (I_0) was in agreement with the size of a dimer. Taking advantage of the available knowledge regarding ATP-binding sites (Hopfner and Tainer 2003) and our overall envelope,

we built a model of the dimer using the RecN*Head* crystal structure (Figure 18). The calculated theoretical scattering curve derived from this model was superimposed on the experimental SAXS curve. The excellent fit with a $\chi 2$ value of 1.56 indicates that this model describes very well the behavior of the protein in solution (Figure 18). In this model, the coiled-coil domains are located on either side of the two RecN*Head* monomers, that come together in a head-to-tail orientation to form two functional ATP-binding pockets as observed previously for dimerisation of NBDs. The Walker-A K67 and Walker-B E472 residues from one molecule form a functional ATP binding pocket with the signature sequence from the second molecule. In the structure of RecN*Head*, the signature sequence is situated only 9 Å away from the Walker-B motif, while this distance increases up to ~20 Å for SMC or Rad50 head domains, suggesting that the two ATP binding pockets formed at the NBD dimerisation interface would be much closer together in the case of RecN compared to SMC or Rad50.

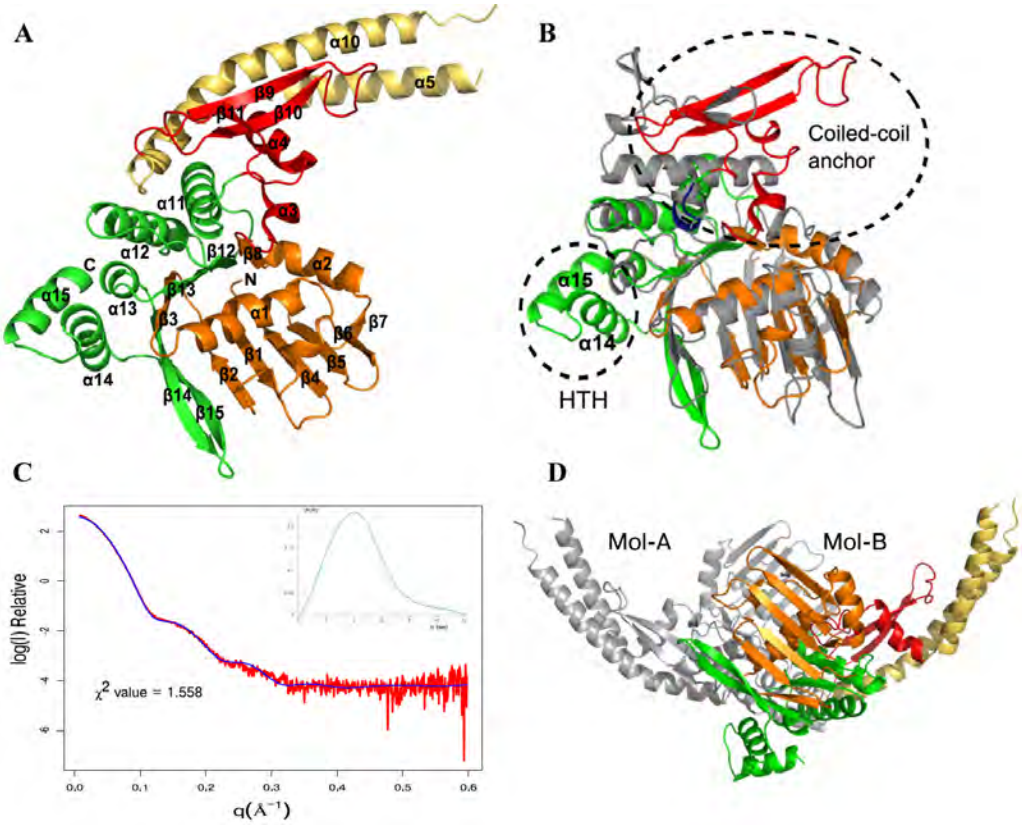


Figure 18: (A) Crystal structure of RecNHead domain. The NBD formed by the N- and C-terminal domains and the beginning of the coiled-coil region are coloured as in Figure 17B. The coiled-coil anchor domain is coloured red. (B) Overlay of RecNHead with the head domain of pfRad50 (grey) displaying a very good fit with root mean square deviations of 1.61Å over 121 Cα. The coiled-coil anchor motif and the C-terminal HTH are indicated by dotted lines. (C) SAXS analysis of RecNHead$^{K67A/E472Q}$. Experimental curve averaged from two protein concentrations (in red). The theoretical curve (in blue) was derived from the model proposed in (D) and the Chi-square value of the fit is reported. In the upper right the pair distribution function (P(r)) relative to the experimental scattering curve is shown. (D) Model of RecNHead$^{K67A/E472Q}$ head-to-tail dimer derived from our SAXS analysis. One molecule is coloured grey, while the second is coloured as in (A).

The central region of RecN proteins is predicted to be a coiled-coil and to be involved in dimer formation (Graumann and Knust 2009). Sequence alignment of *D. radiodurans* RecN with other bacterial RecN proteins highlights how this domain is conserved in length but shows very low sequence identity (< 2%). We determined the crystal structure of RecN*cc*,

a construct corresponding to the coiled-coil domain of RecN (Figures 17B) at 2.0 Å resolution (Table 5). The structure consists of 6 α-helices (α5-α10) of varying lengths, which follow an anti-parallel arrangement (Figure 19A). RecN*cc* forms a stable dimer with a clear two-fold symmetry. The dimer interface is formed by residues located on helices α6-α8 interacting with their equivalent residues from the second monomer through an extended network of van der Waals interactions. The buried interface covers an area of ~1400 $Å^2$ and involves 34 residues. The anti-parallel coiled-coil arrangement of each RecN*cc* monomer is favoured by dipole and hydrogen-bonding interactions (Figure 19A): in particular the highly conserved E222 and R319 form a salt bridge that stabilises the packing of helix α6 against helix α9. As a result, the overall structure of RecN*cc* appears to be very rigid. Several loops located between helices α5-α6, α8-α9 and α9-α10 confer some degrees of freedom to the entire rod-like structure.

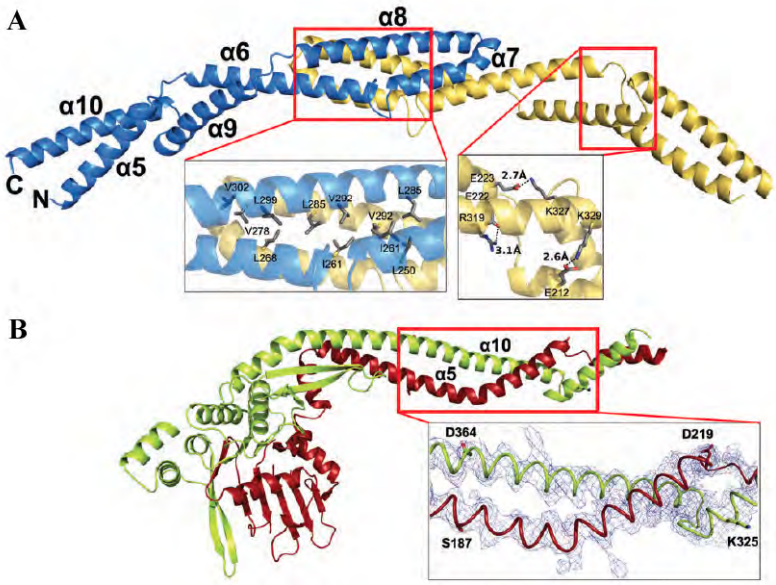


Figure 19: (A) Crystal structure of the dimeric coiled-coil domain of RecN. A close-up view of the dimer interface is shown in the box. Residues involved in dimerisation are represented as sticks. Distances are reported for intermolecular ionic interactions that stabilise the anti-parallel fold of the coiled-coil. (B) Crystal structure of RecNΔdd in which the N-terminal region is coloured red and the C-terminal domain is coloured light green. Helices α5 and α10 are respectively ~75 and ~100 Å long and present large kinks that determine the orientation of the coiled-coil domain. A close-up view of the final electron density map (2Fo-FC map contoured at 1.5σ) is shown for the central region of the coiled-coil, which allowed us to obtain a full model of helices α5 and α10 that were only partially visible in our RecNHead and RecNcc structures.

To confirm whether this dimerisation interface is relevant to dimer assembly in RecN, we characterised a deletion mutant (RecNΔdd) lacking helices α7 and α8. MALLS measurements were performed, confirming that RecNΔdd is a monomer in solution and we determined the crystal structure of RecNΔdd (Figure 19B) at a maximum resolution of 4.0 Å (Table 5). The electron density map allowed us to extend the long and kinked helices α5 and α10 belonging to the coiled-coil domain that were only partially present in the RecN*Head* structure. Coiled-coil domains of SMC and SMC-like proteins display very low sequence conservation but, in contrast, within a protein family are highly conserved in length. SMC proteins, for example, have coiled-coil regions of around 900 amino acids long (Melby, Ciampaglio et al. 1998), whereas bacterial RecN proteins have considerably shorter coiled-

coils composed of only approximately 240 residues. Our crystal structure of the coiled-coil domain of RecN highlights the remarkable diversity of interfaces used for dimer formation among SMC, SMC-like and RecN proteins. Dimer assembly in SMCs and ecMukB involves mostly ionic interactions, whereas Rad50 dimerization is favoured by a zinc-hook motif and a small hydrophobic interface located between the Zn-hook and the coiled-coil (Hopfner, Craig et al. 2002). In RecN proteins, while this region is poorly conserved, the hydrophobic nature of the residues involved in dimer assembly is retained, indicating that this mode of dimerisation is most likely conserved within the RecN protein family.

These three crystal structures of overlapping regions of *dr*RecN were then used to assemble a complete quasi-atomic model of the full-length RecN monomer (Figure 20A). The resulting RecN monomer is very elongated (190 Å long). To analyse the oligomeric state of RecN in solution, we performed SEC coupled to on-line MALLS. RecN elutes as a homogeneous dimer with a molecular mass of 108 kDa in high salt buffer, while it tends to form heterogeneous higher order oligomers in lower salt conditions. SAXS studies were performed on RecN (Figure 20B) and the Rg was estimated to be ~9.5 nm, using the Guinier approximation. The molecular weight estimated from the I_0 value is in agreement with a dimeric assembly of full-length RecN and the profile of the calculated Pair distribution function, characterized by a long tail at higher R values also clearly indicates that full-length RecN is an elongated molecule. A complete model of dimeric RecN (Figure 20C) was prepared using the model of monomeric RecN and the dimeric interface of RecN*cc*. In this model, the two head domains of RecN located ~300 Å apart face each other at the two extremities of the dimer and are twisted 180° apart. The theoretical scattering curve derived from this model showed a very good fit with the experimental SAXS data, giving an estimated $\chi 2$ value of 3.44 (Figure 20B).

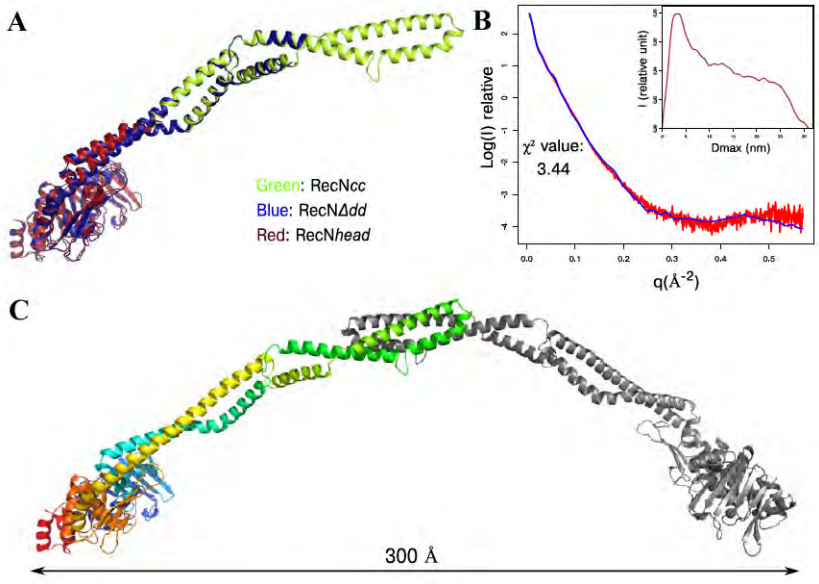
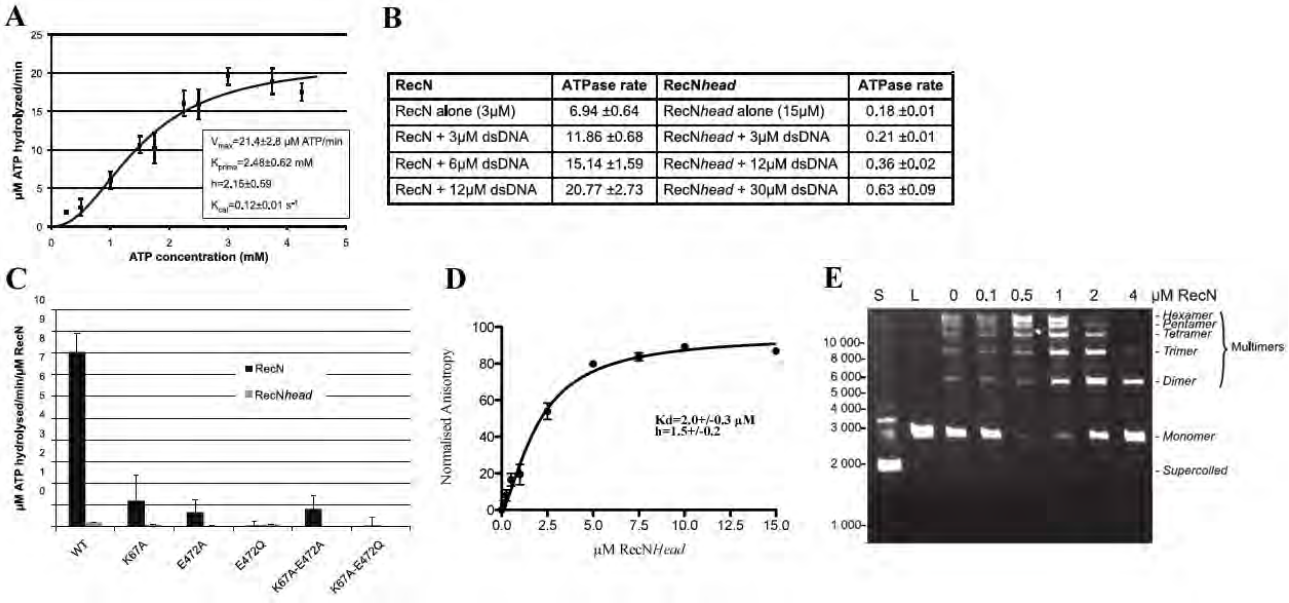


Figure 20: (A) Quasi-atomic model of RecN. (A) Superposition of RecNΔdd (light green) with RecNHead (dark red) and RecNcc (blue) resulted in very good fits with RMSD values of 1.06 Å over 342 Cα for RecNHead and 1.01 Å over 89 Cα for RecNcc domain. (B) SAXS analysis of RecN in high salt conditions. The experimental curve (in red) is superposed with the theoretical curve (in blue) determined from the assembled model represented in (C). The goodness of fit (χ2) is reported. The Pair distribution function is displayed in the upper right square and suggests, from its shape, that the scattering molecule is very elongated (D_{MAX} of 31.5 nm). (C) Architecture of full-length RecN. One molecule of the dimer is coloured in

rainbow, starting from the N-terminus (light blue) to the C-terminus (red), while the other is coloured in grey.

To better relate these structures to RecN's functions, we carried out a number of *in vitro* activity assays on RecN, RecN*Head* and their Walker-A and B mutants (Figure 21). RecN displays a weak ATPase activity with a turnover rate (k_{cat}) of 0.12 s^{-1} (Figure 21A). The enzyme kinetics data were best fit to an allosteric sigmoidal model, indicating that RecN has two cooperative ATP binding sites that do not follow simple Michaelis-Menten kinetics. This is in agreement with previously reported studies on RecN (Reyes, Patidar et al. 2010) and a number of SMC-like proteins (Hopfner, Karcher et al. 2000; Lammens, Schele et al. 2004; Lammens, Bemeleit et al. 2011). A titration of the ATPase activity as a function of protein concentration revealed that the isolated RecN*Head* domain shows a significantly reduced ATPase activity compared to the full-length RecN, suggesting that the coiled-coil region plays an important role in maintaining a basal ATPase activity. Introducing single and double mutations into the Walker-A and -B motifs of RecN resulted in severely reduced (K67A, E472A and K67A/E472A), or completely abolished (E472Q and K67A/E472Q) ATPase activity (Figure 21C), confirming that these residues are directly involved in the ATP binding and/or hydrolysis process in RecN. The activities of Walker-A and -B RecN*Head* mutants were similarly impaired.



RecN	ATPase rate	RecN*head*	ATPase rate
RecN alone (3µM)	6.94 ±0.64	RecN*head* alone (15µM)	0.18 ±0.01
RecN + 3µM dsDNA	11.86 ±0.68	RecN*head* + 3µM dsDNA	0.21 ±0.01
RecN + 6µM dsDNA	15.14 ±1.59	RecN*head* + 12µM dsDNA	0.36 ±0.02
RecN + 12µM dsDNA	20.77 ±2.73	RecN*head* + 30µM dsDNA	0.63 ±0.09

Figure 21: (A) ATPase activity of 3 µM RecN in the presence of increasing concentrations of ATP (0-4.25mM). Kinetic parameters derived from these data are shown. The data was best fit to an allosteric sigmoidal model, indicating that RecN has two (h coefficient) ATP binding sites and does not follow standard Michaelis-Menten kinetics. (B) DNA-stimulated ATPase rates of RecN and RecNHead. The ATPase activity of 3 µM RecN was measured in the presence of 0-12 µM linearized plasmid dsDNA and that of 15 µM RecNHead was measured in the presence of 0-30 µM dsDNA. The ATPase rates are given in µM ATP hydrolysed/min/µM RecN. (C) ATPase activities of 0.25 µM wild-type and mutant RecN (black bars) and of 15 µM wild-type and mutant RecNHead (grey bars). (D) Fluorescence anisotropy measurement of RecNHead binding to fluorescein 5'-end-labelled 50mer dsDNA. (E) DNA end-joining activity of 0-4 µM RecN on blunt-ended linearised pUC19 dsDNA plasmid. After incubation of the plasmid DNA with RecN and then with T4 DNA ligase, the DNA was purified and separated on an agarose gel. As indicated on the right of the gel, RecN favours the formation of plasmid multimers. Control plasmids were also loaded on these gels:

S, supercoiled plasmid and L, linearised plasmid.

ATPase measurements conducted in the presence of increasing amounts of linear plasmid DNA also clearly indicated that the ATPase rates of both RecN and RecN*Head* were stimulated by dsDNA (Figure 21B). Calculation of the electrostatic surface potential of RecN reveals that its surface is largely dominated by negatively charged residues and only one positively charged region, located on the Head domain, was identified as a putative DNA-binding motif. This patch consists of three arginine residues (R81, R120 and R133), two of which are highly conserved in RecN proteins, located on β7, α2 and the loop connecting α1 to β4 of the NBD. These residues are ideally positioned to interact with the negatively charged DNA backbone. Together with our biochemical results these findings indicate that the DNA binding properties of RecN most likely reside in the head domain of RecN. This was confirmed by recent fluorescence polarisation data that showed that isolated RecN*Head* domains could interact with fluorescein end-labelled dsDNA oligonucleotides displaying a K_d of 2.0±0.3 μM (Figure 21D). As with the ATPase activity of RecN, the DNA binding curve of RecN*Head* displays clear cooperativity.

We then investigated the DNA end-joining activity of RecN. As previously reported (Reyes, Patidar et al. 2010), RecN displays a concentration-dependent cohesin-like activity resulting in a clear pattern of plasmid multimerization (Figure 21E). The Walker-A and –B mutants of RecN display mostly reduced DNA-end joining activity compared to wild-type RecN, favouring the accumulation of monomers, dimers and trimers. Nonetheless, it is quite clear that the mutations in the Walker-A and –B motifs have a much greater effect on the ATPase activity of RecN compared to its DNA end-joining activity. The plasmid multimerisation pattern obtained for the RecN mutants is reminiscent of the pattern observed previously for RecN in the presence of a non-hydrolysable analogue of ATP (Reyes, Patidar et al. 2010). These results suggest that the capacity to hydrolyse ATP is not an essential requirement for RecN-driven catalysis of plasmid multimerisation. However, impaired ATPase activity reduces the turnover of RecN and therefore the extent of multimerisation. The isolated RecN*Head* domain, in contrast to the full-length protein, was unable to promote DNA end-joining (Figure 21E). ATP-dependent head-head engagement appears to be essential for the DNA end-joining activity of RecN.

Based on our structural and biochemical data, the head domains are very unlikely to engage to form a closed circular dimeric state as has been proposed for SMCs (Hirano 2002). RecN has a short and rigid coiled-coil, which differs from the high flexibility of the longer coiled-coil domains of SMC proteins (Hirano 2006). Instead, our data suggest a model in which dimeric RecN molecules, which constitute the building blocks, probably form polymers via head-head engagement (Figure 22). Such an assembly would be tightly regulated by the nucleotide-bound state of RecN. Our biochemical data suggests that the interaction of RecN's head domain with dsDNA strongly favours head-head engagement and subsequent polymerisation of RecN along the DNA. We speculate that the assembly of such a helical filament across two broken ends would provide an efficient mechanism to hold these together until the DNA repair machinery is recruited to the sites of DSB. In this model, which has previously been proposed for bacterial SMCs (Hirano 2006), the homologous strands of DNA would be held together by RecN polymers and maintained sufficiently far away from each other by the coiled-coil domain acting as a molecular spacer, in order to avoid unwanted DNA end-joining. A similar mechanism has been proposed for cohesin and Rad50 (Hirano 2002; Lammens, Bemeleit et al. 2011) but our results represent the first structural evidence for such a model. ATP hydrolysis would then be the trigger for disengagement of the head domains and disassembly of RecN from the DNA, in order to recycle RecN for a new round

of DNA tethering and to allow the DNA repair machinery to activate DNA strand exchange and homologous recombination (Figure 22).

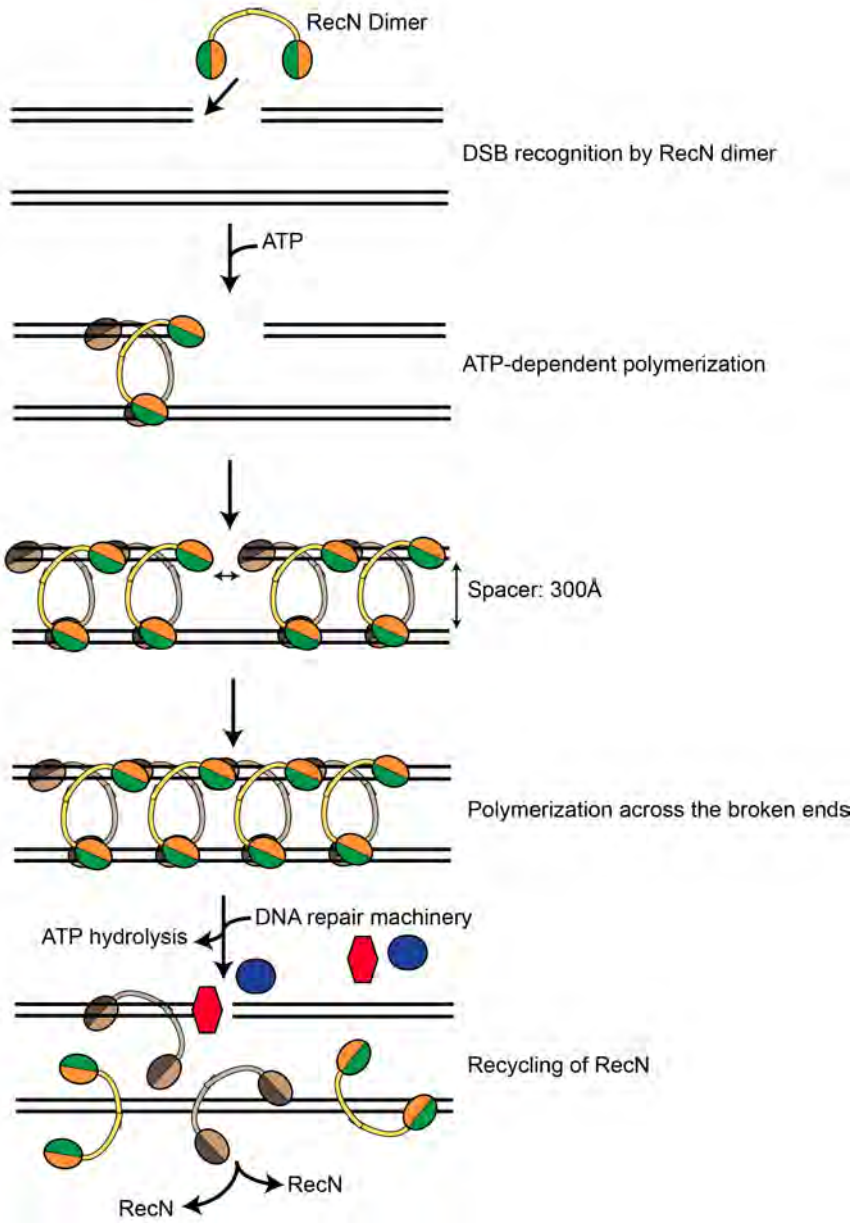


Figure 22: Model for DSB recognition by RecN. DSBs are recognised by RecN in the form of an open, elongated dimer. Recognition and binding to the DNA occurs through the individual head domains of RecN. Assembly on the DNA then favours nucleotide binding and head-head engagement, thus leading to the formation of tetramers and subsequently higher order oligomers as RecN polymerises into a helical filament that wraps around the damaged DNA and its sister chromatid (spatial separation between the two homologous strands corresponds to ~300Å). Head-head engagement of two RecN molecules on either side of the DSB would maintain the DNA ends close together and make them available for homologous recombination. Arrival of the repair machinery may stimulate ATP hydrolysis, which would lead to disengagement of the head domains and disassembly of the RecN filament. The released RecN would then be available for a new round of DSB recognition and homologous recombination could take place.

2. Nucleotide excision repair

In 2004, the ESRF Structural Biology Group decided to initiate a new phase of their structural genomics project, which was more project-based. Twenty targets, including five DNA repair genes belonging to the Uvr (UV resistance) pathway, were selected for this new round of production. After an in-depth bioinformatics analysis of the genes and a careful design of the constructs on our side, the cloning and initial optimisation of expression and purification conditions were again outsourced to Protein'eXpert in Grenoble. The Uvr proteins (UvrA1, UvrA2, UvrB, UvrC and UvrD) were successfully cloned and expressed in *E. coli* and after one year in the hands of Protein'eXpert, the project returned to ESRF and in 2005, together with a 1-year Masters' student (Sofia Caria) I began working on the Uvr proteins from *D. radiodurans*.

2.1. UvrA proteins

UvrA proteins are one of the key components of the nucleotide excision repair (NER) pathway responsible for the repair of the structurally diverse lesions caused by ultra-violet light which involves the recognition and removal of damaged DNA by a dual-incision event (reviewed in (Truglio, Croteau et al. 2006)). In the NER pathway in prokaryotes (Figure 23), it is generally accepted that UvrA, acting as a dimer together with UvrB, is responsible for DNA damage recognition. In addition, UvrA proteins on their own have been shown to bind preferentially to damaged DNA (Mazur and Grossman 1991; Thiagalingam and Grossman 1991). After damage recognition, UvrA dissociates from the DNA, while UvrB forms a stable pre-incision complex upon sites of DNA damage. UvrC subsequently binds to the UvrB-DNA complex and incises the DNA first on the 3' side, then on the 5' side of the lesion (Sancar and Rupp 1983). The resulting 12 to 13 nucleotide fragment containing the damaged DNA is released by UvrD and the gap is then filled by DNA polymerase I and ligase.

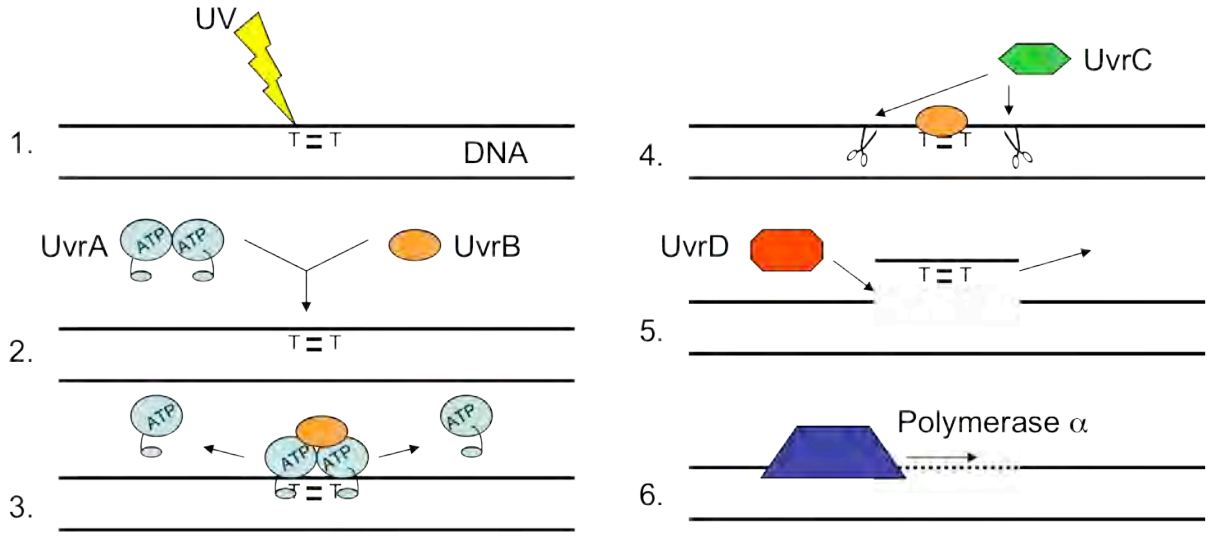


Figure 23: Schematic diagram describing the steps leading to repair of UV-induced DNA damage by the NER pathway in prokaryotes involving the UvrA, UvrB, UvrC and UvrD proteins.

The availability of a large number of bacterial genomes has revealed that the UvrA family consists of five different classes (Goosen and Moolenaar 2008). Of these classes the best studied is the class I UvrA. The primary differences between the various class members

are either deletion of specific domains (class II and class III) or gene duplication (class IV) or both (class V). Class I and class II UvrA genes show a high degree of sequence similarity, where the deletion observed is restricted to the UvrB recognition/interaction domain. *D. radiodurans* encodes two UvrA proteins, the class I *dr*UvrA1 and class II *dr*UvrA2. In *D. radiodurans* the expression of both of these genes is up regulated during the early to mid-phase following irradiation (Liu, Zhou et al. 2003). Despite this lack of a UvrB binding domain, there is evidence that class II UvrAs play a minor role in DNA repair and tolerance to DNA damaging agents, such as UV or chemical treatment (Tanaka, Narumi et al. 2005; Shen, Chiang et al. 2007). The precise role of UvrA2 remains unclear due to the redundancy of the multiple UV-damage repair pathways in bacteria. A recent study of *P. putida* NER proteins indicates that UvrA2 is also involved in repair of damaged DNA, a role that becomes evident only in the absence of UvrA1 (Tark, Tover et al. 2008).

In 2005 when we started to work on *dr*UvrA1 and *dr*UvrA2, it became rapidly apparent that *dr*UvrA2 was more suitable for structural studies, since *dr*UvrA1 had a strong tendency to aggregate. Sofia Caria, a Portuguese Masters' student purified and crystallised *dr*UvrA2 and it took us nearly one year to solve its crystal structure. A construct corresponding to residues 81-922 of *dr*UvrA2 was cloned, expressed and purified with a N-terminal His-tag in the presence of ATP on a Ni-affinity column followed by a SEC step. Two crystal forms were obtained by vapour diffusion with reservoir solutions consisting of 17% PEG 3000, 0.1M citrate pH 5.2, 1mM TCEP +/- 125mM ammonium sulphate. These belonged to the *C2* and $C222_1$ space groups and contained respectively three and two molecules per asymmetric unit. Diffraction data were collected on ID29 and ID23-1 at the ESRF and the structure was initially phased by a three-wavelength multiple-wavelength anomalous dispersion (MAD) experiment around the Zn K absorption edge of a single *C2* crystal of *dr*UvrA2. However, high quality experimental electron density was only obtained by multi-crystal density modification and averaging using data from both crystal forms, which allowed, with the aid of selenomethionine positions obtained from anomalous difference fouriers, an almost complete model of the structure of *dr*UvrA2 to be constructed. The *C2* and $C222_1$ models were refined respectively to 2.3 Å and 3.0 Å resolution with R_{fact}/R_{free} of 21.9/29.4% and 21.7/29.4% (Table 6).

In both crystal forms, the functional unit of *dr*UvrA2 is a dimer with two ADP molecules bound to each monomer. Fortuitously, each copy of the molecule adopts a different conformation, thus providing a view of five distinct monomeric and three dimeric conformational states. The *dr*UvrA2 dimer is a saddle-shaped molecule, consisting of a core region with two elongated domains protruding from this central core. Each monomer of *dr*UvrA2 contains two ABC-type NBDs, an insertion domain (ID) and two zinc-binding modules (Figure 24). Such a domain organisation is commonly found in eukaryotic ABC proteins (Higgins 1992); prokaryotic ABC proteins in contrast usually possess a single NBD. The ID consists of a separate globular domain that exhibits a unique fold, consisting of a core helical domain with two anti-parallel β-strands and an additional α-helix inserted between two semi-continuous helices. Each NBD consists of two lobes. Lobe 1 the catalytic, RecA-like subdomain, contains the Walker A motif and consists of two β-sheets flanked by α-helices. Lobe 2 is an α-helical subdomain and contains the ABC signature sequence. In a monomer the two NBDs (NBD-I, NBD-II) form an intramolecular dimer and are arranged in a head-to-tail fashion. Zn binding module 2 is inserted into the signature sub-domain of NBD-II while Zn binding module 3 is inserted into the signature sub-domain of NBD-I. Two ADP molecules are bound in composite nucleotide binding pockets formed by the Walker A motif of one NBD and the ABC signature sequence of the second. The Walker A motifs, located at the N-terminus of helices α1 and α17 are highly conserved in UvrAs and participate in a

dense network of hydrogen bonds which support the phosphate groups, while the adenine moieties are stabilised by π-stacking interactions. In both binding sites the amino group of the conserved lysine (K49 and K536) and the hydroxyl group of the adjacent serine form hydrogen bonds to the β-phosphate of the ADP.

Table 6. *Summary of crystal parameters, data collection and refinement statistics for drUvrA2. The numbers in parentheses are those for the highest resolution shells.*

Data statistics						
X-ray data	**Zn-SAD**	**High Resolution**	**Zn-MAD**			**Se-SAD**
Space group	*C222₁*	*C2*	*C2*	*C2*	*C2*	*C2*
Unit cell	*a* = 149.7 Å, *b* = 170.9 Å, *c* = 204.3 Å	*a* =269.8 Å, *b* = 110.1 Å, *c* = 101.0 Å, β=100.8°	*a* = 271.6 Å, *b* = 110.9 Å, *c* = 102.7 Å, β=101.0°			*a* =268.7 Å, *b* = 109.1 Å, *c* = 102.4 Å, β=101.0°
Wavelength (Å)	1.2825 (pk)	0.9800	1.2828 (pk)	1.2830 (ip)	0.9756 (re)	0.9793 (pk)
Resolution range (Å)	50.0 - 3.0 (3.2 - 3.0)	47.3-2.3 (2.4 - 2.3)	46.6-2.6 (2.7- 2.6)	46.6-2.6 (2.7- 2.6)	46.9-2.4 (2.5-2.4)	30.0-3.5 (3.7 - 3.5)
R_{sym} (%)	9.7 (54.9)	7.7 (59.2)	5.0 (59.8)	6.3 (80.0)	5.7 (88.7)	10.0 (28.0)
Unique reflections	50,546 (7,444)	123,364 (12,065)	84,618 (8,840)	84,644 (8,707)	75,469 (8,057)	36,763 (5,332)
<(I)/σ(I)>	11.2 (2.1)	7.1 (1.1)	14.1 (1.5)	12.2 (0.9)	9.1 (0.7)	16.6 (4.8)
Completeness (%)	96.4 (98.2)	92.1(62.1)	90.0(50.9)	94.1(67.4)	66.1(48.9)	99.8 (100)

Refinement		
R / R_{free}(%)	21.7/29.4	21.9/29.4
<B> (Å^2)	67.5	57.3
r. m. s. deviations from ideal geometry		
Bond lengths (Å)	0.018	0.014
Bond angles (°)	1.964	1.536
Ramachandran plot		
Favoured (%)	79.7	88.0
Allowed (%)	19.1	11.1
Generously allowed (%)	1.0	0.6
Disallowed (%)	0.1	0.2

The two ADP binding sites observed in each monomer of *dr*UvrA2 are not identical. Such an asymmetry of nucleotide binding sites in dimeric NBDs has been shown to be essential for efficient regulation of protein function (Lamers, Winterwerp et al. 2003; Zaitseva, Oswald et al. 2006). An analysis of all available structures of NBDs indicate that a number of structural rearrangements occur upon nucleotide binding, which has led to the postulation of an induced fit mechanism (Karpowich, Martsinkevich et al. 2001). Notably the Q-loop linking the two subdomains of each NBD appears to act as a switch. The conserved glutamine residue characteristic of this motif is believed to bind the catalytic magnesium ion and/or the nucleophilic water molecule of the ATP hydrolysis reaction (Hopfner and Tainer 2003).

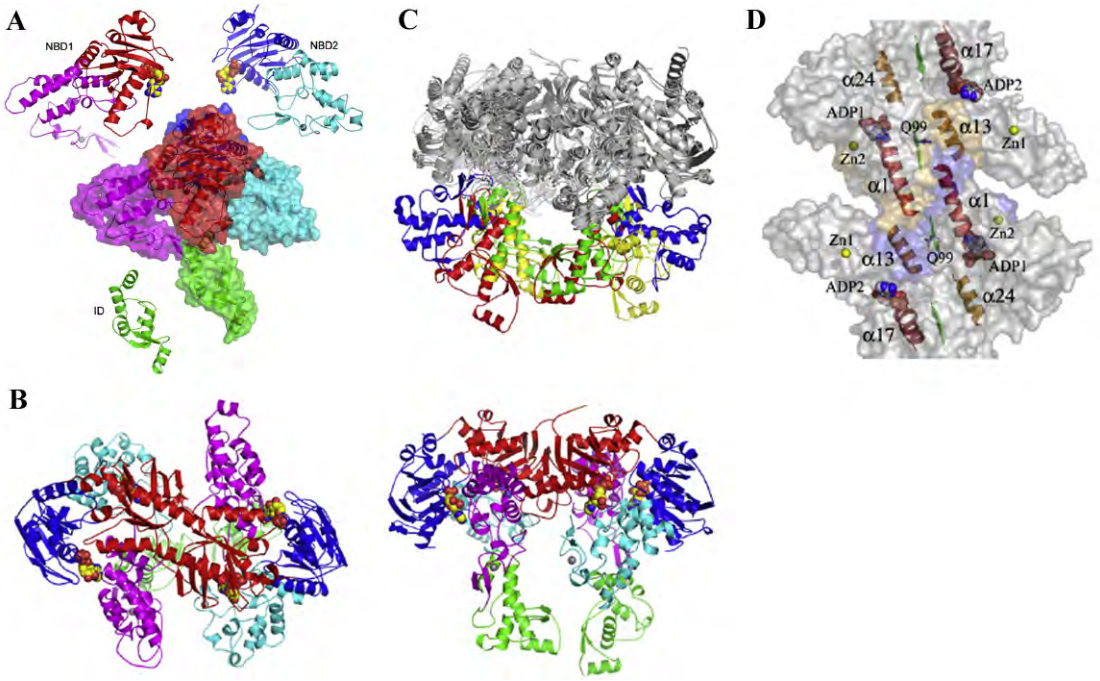


Figure 24: Structure of drUvrA2. (A) Surface and ribbon representation of monomeric drUvrA2. The domains are coloured as follows: NBD-I ATP-binding domain (red), NBD-I Signature domain (magenta), NBD-II ATP-binding domain (blue), NBD-II Signature domain (cyan) and Insertion domain (ID; green). Zinc ions and ADP molecules are illustrated as spheres. (B) Top and side views of the dimeric assemblies of drUvrA2. (C) Superposition of the bstUvrA dimer (blue) and the three drUvrA2 dimers observed in our crystal structures (C2 AB dimer in red, C2 CC dimer in green and $C222_1$ AB dimer in yellow), showing the high degree of flexibility demonstrated by the Insertion Domains (coloured). The core NBDs of the four dimers are shown in grey. (D) Surface representation of drUvrA2 dimer showing the regions involved in dimer formation. The interfacing residues are coloured in pale yellow (monomer A) and light blue (monomer B). The conserved motifs are highlighted: Walker A (red), Walker B (green), signature sequence (orange) and Q-loop (blue).*

A number of observations suggest that in our structures, NBD-II displays a more closed, ATP bound-like conformation, than NBD-I. In particular, overlaying the ATP binding sub-domains of NBD-I and NBD-II of *dr*UvrA2 reveals a rotational movement of ~5° of the signature sub-domain relative to the ATP binding sub-domain in NBD-II. In addition, in our structures the Q-loops adopt different conformations in the two NBDs. Notably, the conserved Gln99 from the Q-loop of the more open NBD-I domain of *dr*UvrA2 points away from the bound nucleotide, whereas in NBD-II, the equivalent residue (Gln591) points inwards towards the β-phosphate of the ADP. Thus upon ATP binding we expect the signature sub-domain to move so that the Q-loop can directly interact with the nucleophilic water involved in ATP hydrolysis, thereby pulling the entire subdomain, and its associated ID, closer to the nucleotide binding pocket (Oswald, Holland et al. 2006).

The dimer interface is formed in the main by contacts between NBD-I subunits (Figure 24B) and the four nucleotide binding sites contained in the dimer are not at the dimer interface. In 2008, the crystal structure of *Geobacillus stearothermophilus* UvrA1 (*bst*UvrA) was reported (Pakotiprapha, Inuzuka et al. 2008). While the crystal structure of *dr*UvrA2 monomer is very similar to that seen for *bst*UvrA (rmsd of 1.24Å for 593 aligned Cα atoms), on closer inspection, however, there is an interesting difference in the structures of *bst*UvrA and *dr*UvrA2 dimers. In the former, the two IDs contained are well separated (~40 Å apart)

and allow unhindered access to regions of the ventral surface of UvrA, which have been shown to be involved in the binding of DNA (Croteau, DellaVecchia et al. 2008; Pakotiprapha, Inuzuka et al. 2008). In contrast, in the *dr*UvrA2 dimer, the IDs are much closer together and the underside of the core is no longer exposed to allow the direct binding of DNA. While the two conformations observed are probably stabilised by their respective crystal packing, it is clear that UvrA dimers can exhibit conformational variability and can adopt 'ID open' (seen in the crystal structure of *bst*UvrA) or 'ID closed' (seen in the crystal structures of *dr*UvrA2) configurations (Figure 24C). Further evidence of this conformational variability is seen when comparing the structures of the three dimers we see in the two crystal forms in which we have solved the structure of *dr*UvrA2. In each case the orientation of the IDs relative to the dimer core is different and the configuration of the IDs more or less closed.

Although the nucleotide binding sites in *dr*UvrA2 and *bst*UvrA are not located at the dimer interface, they are, however, adjacent to the structural elements involved in the stabilisation of the dimer (Figure 24D). The close proximity between the dimerisation interface and the nucleotide binding pockets ensures that the formation of the dimer is tightly regulated by the nucleotide binding state of UvrAs. Elongation factor eEF3 also contains two NBDs per monomer and in the absence of nucleotide, NBD-II is rotated by 120° away from NBD-I in the elongation factor monomers (Andersen, Becker et al. 2006). In the presence of ATP the two NDBs adopt the tandem formation seen in nucleotide-bound homodimers of MJ0796 (Smith, Karpowich et al. 2002) and Rad50 (Hopfner, Karcher et al. 2000), in nucleotide-bound monomers of *bst*UvrA and in the structures of *dr*UvrA2 monomers. Thus, in the ABC protein family nucleotide binding favours the tight association of two NBDs. This nucleotide-dependent association of the two NBDs of *dr*UvrA2 into a compact structure most likely provides a suitable interface for the head-to-head dimerisation of *dr*UvrA2. In the absence of nucleotide it is therefore probable the two NBDs of UvrAs, which are connected by a long linker region, adopt an open conformation, as seen for nucleotide-free eEF3. This would account for the increased instability of nucleotide-free *dr*UvrA2 we observe in solution. In addition, in one of the dimers we observe, the conserved Gln99, located in the Q-loop, contributes to the stability of the dimer when it is pointing away from the nucleotide binding pocket. Flipping of this residue upon ATP binding and hydrolysis (as mentioned previously) may act as a switch mechanism directly coupling monomer-dimer transition to ATP binding and hydrolysis.

DNA binding experiments were carried out with *dr*UvrA2. *dr*UvrA2 can bind to both undamaged dsDNA and to dsDNA containing a fluorescein-adducted thymine (FldT), a known substrate for the UvrABC excinuclease (DellaVecchia, Croteau et al. 2004; Van Houten, Croteau et al. 2005). Quantification of the amount of free DNA also reveals that *dr*UvrA2 binds preferentially to dsDNA containing a lesion (Figure 25). The location of the damage within the oligonucleotide also seems to be important since *dr*UvrA2 binds preferentially to duplexed DNA with the FldT-modified base located in the middle of the 32-mer, suggesting that *dr*UvrA2 requires duplexed DNA on either side of the lesion for stable association on the DNA. A surface representation of UvrA dimers reveals a long positively charged groove that runs diagonally across the ventral surface of UvrAs (Figure 26), which could accommodate a 32-mer dsDNA. Interestingly, in the 'ID open' conformation of UvrA dimers observed in the *bst*UvrA structure, this surface is freely accessible to DNA; this is not the case in our *dr*UvrA2 structures. Thus, in order for DNA to interact with the ventral surface of *dr*UvrA2 dimers the two IDs must either (a) move apart to allow DNA access to this surface (i.e. they are not directly involved in DNA binding) or (b) clasp onto the DNA and transport it to the ventral surface of the dimer (i.e. the IDs are directly involved in DNA binding).

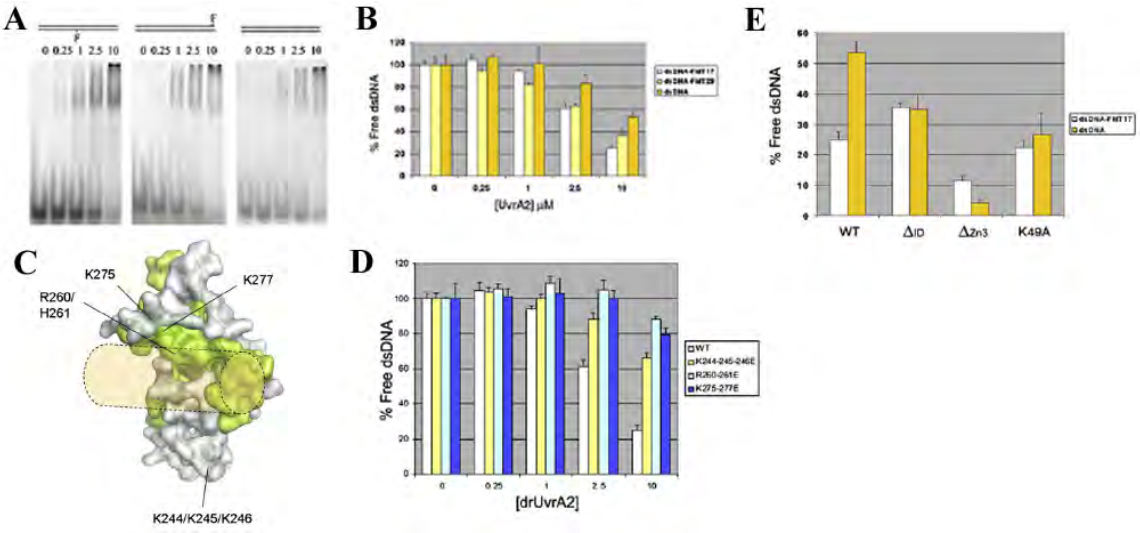


Figure 25: (A) EMSAs of drUvrA2 binding to intact and fluorescein-adducted DNA. Prior to electrophoresis, increasing amounts of wild-type drUvrA2 (0, 0.25, 1, 2.5 and 10 µM) were incubated with 1 µM of intact (orange) or fluorescein-adducted thymine (at position 17 or 29; respectively white and yellow) containing dsDNA. (B) The free DNA in the EMSAs (A) was quantified, averaged from three measurements in each case and plotted as histograms. (C) Surface representation of the ID of drUvrA2 showing the strictly conserved residues in dark green and highly conserved residues (≥70% conservation in UvrAs) in pale green. The mutated residues are indicated and an orange cylinder indicates the proposed position of the bound dsDNA. (D) dsDNA binding activity of wild-type (WT) and mutant drUvrA2s. Increasing amounts of drUvrA2 (0, 0.25, 1, 2.5 and 10 µM) were incubated with 1 µM of fluorescein-adducted thymine (position 17) containing dsDNA prior to electrophoresis. The free dsDNA was quantified, averaged from three measurements in each case and plotted as histograms. (E) Histogram illustrating the differential binding of 10 µM WT, Δ_{ID}, Δ_{Zn3} and K49A mutant drUvrA2 to 1 µM intact (orange) and damaged (white) dsDNA.

A number of observations suggest that the IDs are directly involved in DNA binding. Calculation of the electrostatic surface potential of *dr*UvrA2 and *bst*UvrA reveals conserved positively charged patches on the insides of the IDs and in the case of *dr*UvrA2 also at the tips of the IDs. These regions could provide interaction surfaces for negatively charged DNA. Moreover, primary sequence analysis of UvrA proteins also reveals that the IDs contain a significantly higher number of conserved positively charged and aromatic residues compared to the rest of the protein. Also a search for structural homologues of the *dr*UvrA2 ID identified a ribosomal protein chain M (Tung, Joseph et al. 2002), which is similar in structure to the central helical core of the domain (rmsd of 1.8 Å for 40 aligned Cα atoms). Interestingly, ribosomal protein M interacts with duplex RNA and suggests a possible mode of how the IDs in UvrAs might interact with DNA. The resulting model of *dr*UvrA2 bound to dsDNA (Figure 26) based on this is reminiscent of the MutS-DNA complex (Lamers, Perrakis et al. 2000), which also involves a core NBD domain linked to a more flexible DNA-binding domain.

To test whether the IDs are directly involved in DNA binding, we prepared a number of point mutants in order to remove positive charges from regions of the IDs that, according to our model, could be involved in DNA binding. Two of these regions are located in the conserved central helical core of the ID (residues R260, H261, K275 and R277 in helices α8 and α9), while a third is situated on the helix forming the tip of the domain (residues K244,

K245 and K246). This helix is missing in the structure of *bst*UvrA and is not conserved within the UvrA family. For the two mutants, R260E/H261E and K275E/R277E, binding to DNA is severely disrupted (Figure 25D), indicating that the central region of the IDs is essential for dsDNA binding. The triple mutant (K244E/K245E/K246E) displays a less severe phenotype compared to the R260E/H261E and K275E/R277E mutants – its ability to bind to dsDNA is nonetheless reduced compared to wild-type *dr*UvrA2, indicating that this region may also be playing a role in dsDNA binding. One possibility is that this positively charged region allows UvrA to recruit and guide the DNA towards the inside of the IDs.

Two mutants of *dr*UvrA2: one in which the entire ID was deleted (*dr*UvrA2-Δ_{ID}) and one in which the C-terminal zinc-finger motif associated was truncated from CXXC–X_{19}–CXXC to CXXC–X_5–CXXC (*dr*UvrA2-Δ_{Zn3}) were assayed for their ability to bind dsDNA with or without a fluorescein-adduct (Figure 25E). Both mutants exhibited significant binding to both types of dsDNA substrates, thus confirming previous studies, which revealed that the ID and the C-terminal zinc-finger motif of UvrAs were dispensable for DNA binding (Croteau, DellaVecchia et al. 2008; Pakotiprapha, Inuzuka et al. 2008). That *dr*UvrA2-Δ_{ID} binds dsDNA more efficiently than wild-type *dr*UvrA2 is perhaps counter-intuitive as we also show that residues within the ID are crucial for dsDNA binding by intact *dr*UvrA2. However, in both our deletion mutants (*dr*UvrA2-Δ_{ID} and *dr*UvrA2-Δ_{Zn3}), the ventral surface of the dimer would be more freely accessible to DNA and our results confirm previous observations that the ventral surface of the UvrA dimers is also involved in DNA binding.

However, unlike wild-type *dr*UvrA2, both our deletion mutants display a complete loss in the ability to differentiate between intact and damaged DNA (Figure 25E). This effect can be observed principally as a large increase in binding of the deletion mutants to non-damaged dsDNA compared to wild-type *dr*UvrA2. Removal of the tip of the C-terminal zinc-finger motif from *B. caldotenax* UvrA was shown to result in the same dsDNA binding profile, suggesting that this region is indeed critical for damage recognition (Croteau, DellaVecchia et al. 2006). Our data clearly indicates that efficient recognition of DNA lesions requires the presence of both the ID and the C-terminal zinc-finger.

The process leading to damage recognition is known to be regulated by the nucleotide-bound state of UvrAs (Thiagalingam and Grossman 1991; Thiagalingam and Grossman 1993). Both the ID and the C-terminal zinc-finger are inserted within the NBDs. Conformational changes associated with nucleotide binding and hydrolysis, such as movement of the Q-loop, would thus be transmitted directly to these domains. To further investigate the role of nucleotide binding and hydrolysis on damage recognition by *dr*UvrA2, we therefore studied the ability of the Walker A lysine mutant K49A to bind to our two dsDNA substrates. Again, as with the domain-deletion mutants described above, dsDNA binding is retained but the K49A mutant protein exhibits a significantly impaired capacity to differentiate between intact and damaged dsDNA (Figure 25E), suggesting that an intact ATPase activity is also a requirement for efficient recognition of DNA lesions.

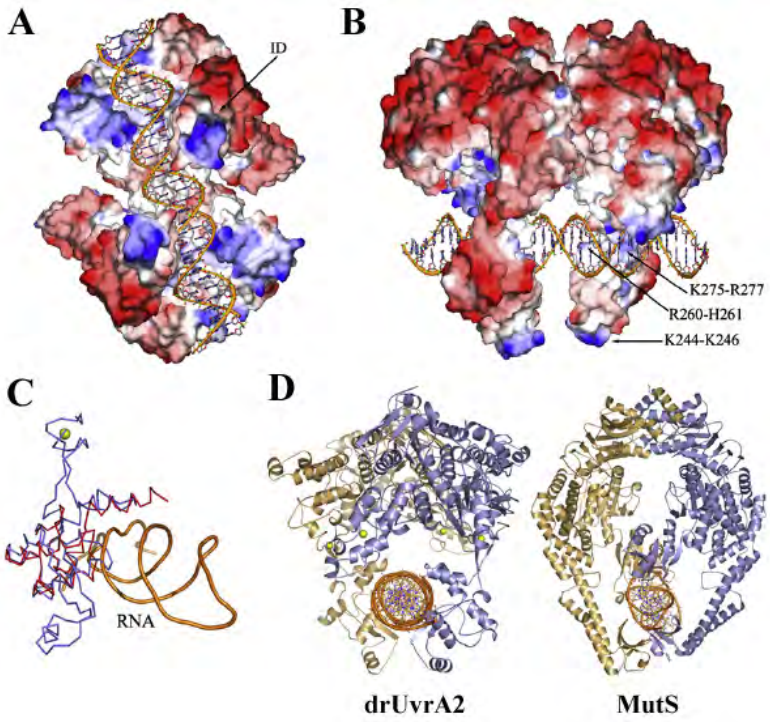


Figure 26: Proposed model for DNA binding. (A)-(B) Illustrations of the electrostatic surface potential of drUvrA2 (symmetric dimer of C2-Mol B; semi-open conformation of IDs) contoured at ±5 kT/e, where red describes a negative and blue a positive potential. The modeled DNA is shown in orange. The residues mutated in our study are indicated by arrows. (C) Superposition of drUvrA2 ID (blue) and ribosomal protein (pdb: 1m5g) chain M (red) bound to RNA (orange). (D) Model of dimeric drUvrA2 bound to dsDNA showing the similarity to the crystal structure of MutS bound to mismatched DNA. For both drUvrA2 and MutS, one monomer is colored in pale yellow and the other in light blue.

Our mutational and structural analyses of *dr*UvrA2 unequivocally assign a significant new role for the Insertion Domain of UvrAs in DNA binding and damage recognition. Our results allowed us to propose an improved model for DNA binding and damage recognition by UvrAs and represent a key contribution to our understanding of the initial steps in bacterial NER. They also pave the way for further structure-guided experiments to further dissect the detailed molecular mechanisms of UvrA action. This work was published in Structure in 2009 (Timmins, Gordon et al. 2009).

2.2. UvrA1, UvrB and UvrC proteins

The UvrB gene was also cloned and successfully expressed in *E. coli* by Protein'eXpert. The protein was highly soluble and could be purified in large amounts. At ESRF, two Masters' students (S. Caria and T. Johansson) under my supervision and a post-doctoral fellow (O. Sleator) worked on the crystallisation of UvrB alone and in complex with DNA, but no crystals were obtained. In 2008, T. Klar, a post-doctoral fellow was recruited to study the UvrA-UvrB interaction. T. Klar optimised the expression and purification protocol for UvrA1 and also cloned, expressed and purified the domains of UvrA1 and UvrB expected to be involved in protein-protein interactions. But unlike their homologues from *G. stearothermophilus* or *E. coli*, *D. radiodurans* UvrA (both UvrA1 and UvrA2) and UvrB proteins do not form a stable complex suitable for structural studies.

In 2010, together with Samira Acajjaoui, the laboratory technician from the ESRF Structural Biology Group and Mats Ökvist, a scientist in the group, I started to work on *D. radiodurans* UvrC, which according to Protein'eXpert's work was mostly insoluble. The full-length gene and a construct corresponding to the C-terminal domain (residues 366-617) were cloned into a bacterial expression vector, expressed in *E. coli* and purified to homogeneity. The intact UvrC required high salt for stability. In such conditions, it appears to exist in the form of both dimers and monomers in solution as evaluated by SEC, MALLS and SAXS. So far, no crystals were obtained for UvrC. The C-terminal domain was purified in large amounts and crystals were obtained using the EMBL crystallisation robot in conditions containing 25% PEG 5,000 MME or 25% PEG 3,350 and 0.1M Bis-Tris pH6.5. Diffraction data was collected on these thin plates on ID14-4 to a maximum resolution of 1.8 Å (Table 7).

Table 7 *Crystallographic data and refinement statistics for UvrC-Cterm. Values in parentheses are for highest resolution shell.*

Data statistics	**UvrC-Cterm**
Space group Unit cell (Å/º)	$P2_1$ a=50.4, b=40.8, c=59.6, β=90º
Resolution range	40.81-1.80 (1.90-1.80)
Rsym (%)	6.5 (31.0)
<I>/<σI>	13.7 (4.0)
Completeness	99.6 (98.2)
Twin fraction	0.3
Refinement	
Unique reflections	21,462
Rfact/Rfree	16.6/20.6
Mol/asu	1
Average B-factor	25.4
Rms deviations Bonds (Å) Angles (º)	 0.020 1.846

The crystal structure was solved by molecular replacement using Phaser and the two individual domains composing the crystal structure of the C-terminal region of *T. maritima* UvrC (Karakas, Truglio et al. 2007) as search models. The C-terminal region of *D. radiodurans* UvrC is composed of the 5' RNase H endonuclease domain and two Helix-hairpin-Helix (HhH) motifs forming a $(HhH)_2$ domain (Figure 27). The structures of the individual domains are very similar to those of *T. maritima* UvrC, but the relative orientation of the HhH motifs with respect to the endonuclease domain is quite different. Preliminary SAXS studies on this construct also indicate that the conformation observed in the crystals is different from that adopted in solution, where a more elongated arrangement appears to be seen. New constructs of UvrC are now in preparation to obtain more structural information regarding the central and N-terminal region of UvrC. A manuscript describing these results is in preparation (Ökvist *et al*, 2012).

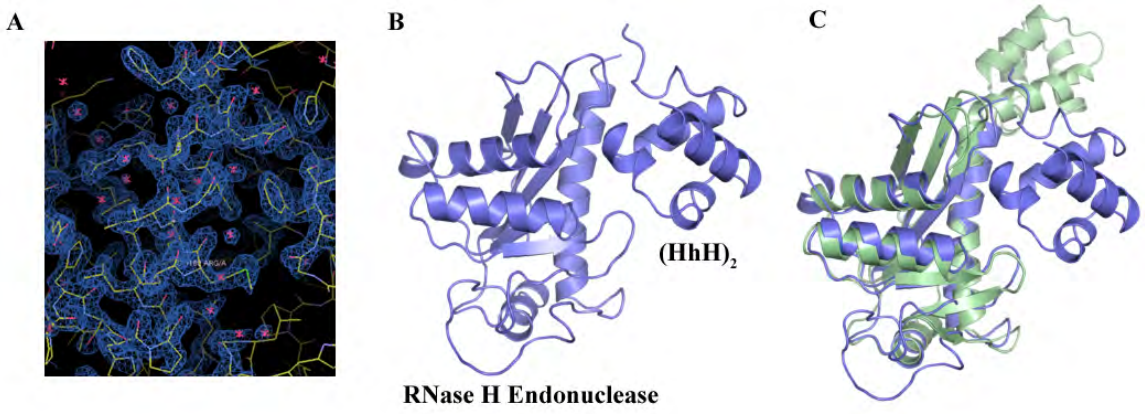


Figure 27: (A) Example of the $2F_o$-F_c electron density maps (contoured at 1σ) obtained for UvrC-Cterm. (B) Crystal structure of UvrC-Cterm. It is composed of two domains: a RNase H endonuclease domain and two HhH motifs forming a $(HhH)_2$ domain. (C) Overlay of the C-terminal regions of D. radiodurans UvrC (blue) and T. maritima UvrC (green).

2.3. UvrD Helicase

In 2008, Samira Acajjaoui started to work under my supervision on the UvrD helicase from *D. radiodurans*, an essential DNA helicase involved in many DNA repair pathways. In 2010, Meike Stelter, a post-doctoral fellow at ESRF (and now working with me at IBS) joined the UvrD project.

Many biological processes, such as DNA replication, transcription, recombination or repair, require access to the genetic information hidden within the duplex DNA of the genome and for this purpose dsDNA needs to be transiently unwound. A diverse set of enzymes, known as DNA helicases, is responsible for catalysing this process (Tuteja and Tuteja 2004b; Tuteja and Tuteja 2004a). DNA helicases are ubiquitous enzymes and many different helicases are found in a single cell due to the diversity of structures adopted by duplexed DNA. Helicases are a subset of the translocase enzyme family that share a number of conserved signature motifs responsible for either NTP binding and hydrolysis, DNA binding or for coupling these two processes. Based on primary structure analyses and extensive biochemical studies, six superfamilies of helicases have so far been described, each of which possesses a different set of conserved signature motifs (Gorbalenya and Koonin 1993; Singleton, Dillingham et al. 2007). Three of these superfamilies (SF1, SF2 and SF6) have been further classified according to their directionality 3'-5' (type A) or 5'-3' (type B) (Singleton, Dillingham et al. 2007).

UvrD is a SF1A helicase (Gorbalenya and Koonin 1993) that plays important functions in DNA replication (Bruand and Ehrlich 2000), recombinational repair (Arthur and Lloyd 1980; Veaute, Delmas et al. 2005; Bentchikou, Servant et al. 2010), methyl-directed mismatch repair (Matson and Robertson 2006) and nucleotide excision repair (Caron, Kushner et al. 1985). UvrD consists of two RecA-like domains (1A and 2A) that are responsible for nucleotide binding and hydrolysis and two additional domains (1B and 2B) that are involved in dsDNA binding. The crystal structures of several UvrD-like helicases in complex with DNA (Korolev, Hsieh et al. 1997; Velankar, Soultanas et al. 1999; Lee and Yang 2006), together with kinetic and single-particle studies (Ali and Lohman 1997; Dillingham, Wigley et al. 2002; Dessinges, Lionnet et al. 2004; Fischer, Maluf et al. 2004) led to the proposal of a combined wrench-and-inchworm mechanism for DNA unwinding (Figure 28A) (Lee and Yang 2006; Yang 2010). In this model, a rotational movement regulated by

ATP binding and hydrolysis acting as the 'engine' is combined with alternate tight and loose interactions at four protein-DNA contact points to produce a highly coordinated unidirectional movement along DNA. An alternative mode of unwinding, known as strand-displacement, has also been reported, notably in the absence of domain 2B and duplex DNA binding (Cheng, Brendza et al. 2002; Lee and Yang 2006), in which movement of ssDNA is deregulated due to reduced contacts with dsDNA.

In *D. radiodurans*, unlike in *E. coli*, UvrD is involved in diverse DNA repair pathways (Bentchikou, Servant et al. 2010). In particular, UvrD has been shown to play a central role in double-strand break (DSB) repair and reconstitution of the genome following chromosome fractionation (Bentchikou, Servant et al. 2010). In *E. coli*, the RecQ, RecD and Helicase IV enzymes participate in DSB repair while in *D. radiodurans*, these three helicases have been shown to be dispensable (Bentchikou, Servant et al. 2010). Interestingly, Helicase IV from *D. radiodurans* exhibits 5'-3' activity, while its *E. coli* homologue unwinds DNA with 3'-5' polarity (Wood and Matson 1987; Cao and Julin 2009), indicating that the involvement, importance and polarity of a helicase in a given cellular pathway are not conserved from one bacterium to another.

Between 2009 and 2011, we determined the crystal structures of *D. radiodurans* UvrD (*dr*UvrD) in complex with 3'-tailed dsDNA in apo- and AMP-PNP bound states and characterised its *in vitro* DNA unwinding activity. The manuscript describing this work has been submitted to Plos One (Acajjaoui *et al*, 2012).

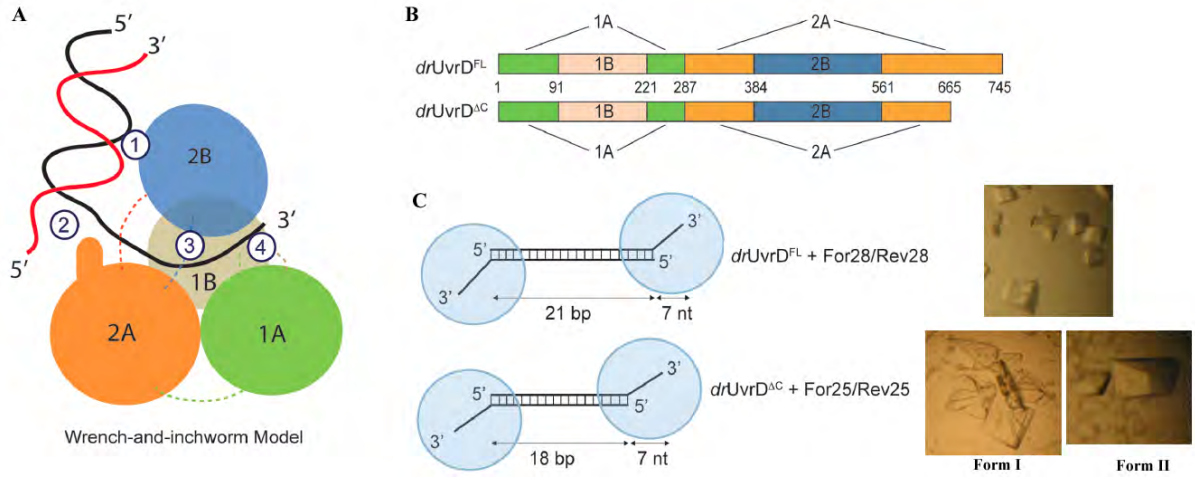


Figure 28: (A) Schematic representation of the wrench-and-inchworm unwinding mechanism, indicating the four critical protein-DNA contact points: the GIG motif interacts with dsDNA (1), the β-hairpin motif with the ss-dsDNA junction (2), motif III with the ssDNA (3) and the ssDNA gateway with the exiting ssDNA (4). (B) Schematic representation of the domain structures of drUvrD^{FL} and drUvrD$^{\Delta C}$. (C) Illustration of the DNA oligonucleotides used for crystallisation of the ternary complexes, containing either drUvrD^{FL} or drUvrD$^{\Delta C}$ bound to 3'-tailed dsDNA and AMPPN. Examples of crystals obtained for each of these complexes are shown on the right.

Full-length *dr*UvrD (*dr*UvrD^{FL}) and a truncated *dr*UvrD (*dr*UvrD$^{\Delta C}$), missing residues 666-745 (Figure 28B), were cloned into a bacterial expression vector and expressed in *E. coli*. The proteins were purified on a Ni-affinity column followed by a heparin-affinity chromatography, yielding milligrams of pure, homogeneous sample. EMSAs were used to identify the optimal DNA substrates (in terms of length of the double- and single-stranded

regions) for co-crystallisation with UvrD, using the work on *E. coli* UvrD as a guide (Lee and Yang 2006). The first crystals were obtained in 2009 with the truncated *dr*UvrD (*dr*UvrD$^{\Delta C}$) in complex with an 18 base-pair dsDNA oligonucleotide with 7 nucleotide extensions at the 3' end and 1 mM AMP-PNP (Figure 28C) using the hanging-drop vapour diffusion method. Two crystal forms were obtained: Form I crystals appeared rapidly (<1day) in 20% PEG 3,350, 0.1M Bis-Tris Propane pH 7.0 and 0.2M Na-Nitrate, while Form II crystals appeared after at least one week in 22% PEG 3,350, 0.1M Bis-Tris Propane pH 7.5 and 0.1M Na-Fluoride. At a later stage of the project, in 2010, we used our experience with *dr*UvrD$^{\Delta C}$ to crystallise the full-length *dr*UvrD with a 21 base-pair dsDNA oligonucleotide with 7 nucleotide extensions at the 3' end. At first, very small, thin plates were obtained in conditions similar to those used for crystals of Form I, and these crystals were used for seeding experiments in order to grow crystals to a suitable size and quality for data collection (Figure 28C). Diffraction data were collected at the ESRF on beamlines ID14-2, ID14-4 and ID23-1 (Table 8). The structure of *dr*UvrD$^{\Delta C}$-For25/Rev25 form I was solved by molecular replacement using the *gs*PcrA helicase as a search model (PDB: 3PJR). After several rounds of substantial rebuilding of the protein chains, the DNA could be built and the AMPPNP molecules docked into the electron density. Subsequently, this model was used to solve the structures of *dr*UvrD$^{\Delta C}$-For25/Rev25 form II and *dr*UvrD^{FL}-For28/Rev28 by molecular replacement with Phaser.

Table 8 *Crystallographic data collection and refinement statistics. Values in parentheses are for highest resolution shell.*

Dataset	***dr*UvrD^{FL}**	***dr*UvrD$^{\Delta C}$ form I**	***dr*UvrD$^{\Delta C}$ form II**
Data collection			
Protein	*dr*UvrD^{FL}	*dr*UvrD$^{\Delta C}$	*dr*UvrD$^{\Delta C}$
DNA	For28/Rev28	For25/Rev25	For25/Rev25
Nucleotide	AMPPNP	AMPPNP	AMPPNP
Space group	$P2_1$	$P2_12_12_1$	$P2_1$
Cell dimensions			
a, *b*, *c* (Å)	71.58, 390.58, 71.65	67.57, 67.45, 386.04	68.49, 89.79, 293.80
α, β, γ (°)	90.00, 106.00, 90.00	90.00, 90.00, 90.00	90.00, 89.97, 90.00
Beamline	ESRF ID14-4	ESRF ID14-2	ESRF ID23-1
Resolution (Å)	46.15 – 4.00 (4.22– 4.00)	47.40 – 2.55 (2.69 – 2.55)	47.63 – 3.00 (3.16 – 3.00)
R_{merge} (%)	10.6 (65.8)	7.1 (59.2)	6.4 (32.0)
<(I)/σ(I)>	10.1 (2.3)	20.4 (3.6)	6.5 (1.9)
Completeness (%)	99.6 (99.5)	100.0 (100.0)	89.0 (86.5)
Refinement			
N° of reflections	30,051	56,088	59,507
R_{fact}/R_{free} (%)	24.6/27.1	21.1/26.6	22.8/28.8
Mol/asu	4 chains UvrD 2 chains dsDNA	2 chains UvrD 1 chains dsDNA	4 chains UvrD 2 chain dsDNA
Ligands	4 AMP-PNP	2 AMP-PNP	2 AMP-PNP
Average B-factor (Å^2)			
Protein	203.0	63.9	102.2
DNA	256.8	163.9	132.0
AMPPNP	168.8	29.7	84.6
Solvent	N/A	38.5	84.3
Rms deviations			
Bonds (Å)	0.006	0.017	0.012
Angles (°)	1.1	1.7	1.5

Crystals of *dr*UvrD^{FL}-For28/Rev28 diffracted X-rays to 4.0 Å (Table 8). Despite being present in the crystallised protein, residues 663-745 corresponding to the variable C-terminal region could not be traced in the electron density map, confirming that this region is particularly flexible (Manelyte, Guy et al. 2009). Higher resolution diffraction was obtained from crystals containing *dr*UvrD$^{\Delta C}$. Crystals of form I diffracted to 2.5 Å, while crystals of form II diffracted to a maximum resolution of 3.0 Å. In all structures, each *dr*UvrD monomer is bound to the ds-ssDNA junction at either end of the DNA duplex, thus forming an assembly of one DNA duplex with two UvrD monomers (Figure 29B). As in previous structures of UvrD-like helicases (Korolev, Hsieh et al. 1997; Velankar, Soultanas et al. 1999; Lee and Yang 2006), there are no direct contacts between the two UvrD monomers loaded on a given DNA duplex. In the structures of *dr*UvrD^{FL} and *dr*UvrD$^{\Delta C}$ form I, each of the protein monomers contains one bound AMPPNP molecule, whereas in *dr*UvrD$^{\Delta C}$ form II each assembly is composed of a DNA duplex with an AMPPNP-bound UvrD on one end and an apo-UvrD on the other.

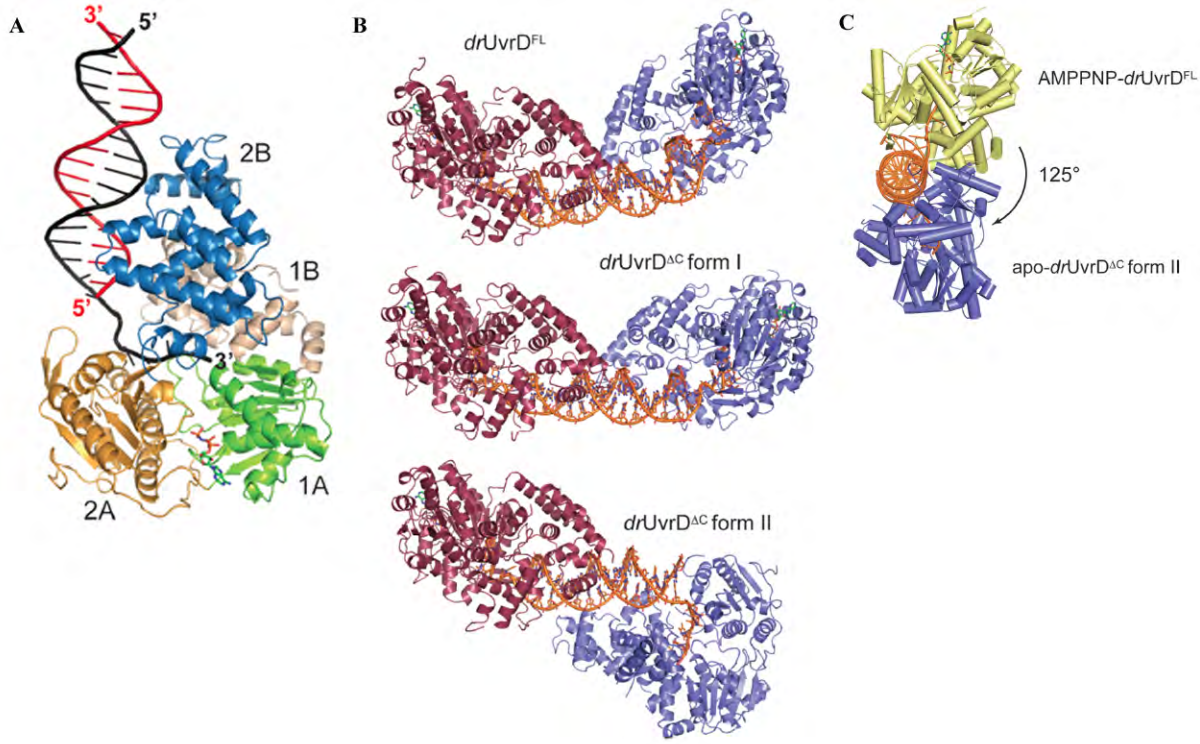


Figure 29: (A) Crystal structure of one monomer of drUvrD$^{\Delta C}$ bound to duplex DNA with a single-stranded extension at the 3'-end. The translocating strand is coloured black and the complementary strand is coloured red. The domains of drUvrD$^{\Delta C}$ are shown in ribbon and are coloured green (1A), beige (1B), orange (2A) and blue (2B). AMPPNP is shown in sticks. (B) Ribbon illustrations of AMPPNP-bound drUvrD^{FL}, AMPPNP-bound drUvrD$^{\Delta C}$ form I and the mixed AMPPNP-bound (red) and apo- (blue) drUvrD$^{\Delta C}$ form II. The DNA and AMPPNP are shown in sticks. (C) Overlay of chains A (red) of drUvrD^{FL} and apo-drUvrD$^{\Delta C}$ form II, illustrating the large spiral movement of chains B coloured respectively yellow, grey and blue. The DNA is shown as an orange ribbon.

*dr*UvrD displays 36% sequence identity with *E. coli* Rep (*ec*Rep) and UvrD (*ec*UvrD) helicases and 42% sequence identity with *G. stearothermophilus* PcrA (*gs*PcrA) helicase, all of which are members of the SF1A helicase family. The overall structures of the *dr*UvrD monomers are very similar to those observed in the closed conformation of *gs*PcrA-DNA and *ec*UvrD-DNA complexes and consist of four domains: 1A, 1B, 2A and 2B (Figure 29A).

When present, AMPPNP is bound at the interface between domains 1A and 2A. ssDNA interacts with all four domains, a majority of contacts being with domain 2A, and interactions with dsDNA involve mainly domain 2B. A close look at domain interfaces reveals that most of the non-conserved residues are found in these areas. The nature of contacts and interface areas between domains 1A/1B and 1B/2B are indeed very different in *ec*- and *dr*UvrD. In *dr*UvrD, these interfaces are significantly smaller than in *ec*UvrD and involve many more ionic interactions. Such differences may impart increased plasticity and flexibility to *dr*UvrD, but may also increase its sensitivity to stress-related changes in its local environment (*e.g.* pH, temperature, salt concentration).

The relative orientation of the components of the protein-DNA assembly observed in *dr*UvrD$^{\Delta C}$ form II is similar to those observed in previous structures of SF1A helicases. Our structures of *dr*UvrD^{FL} and *dr*UvrD$^{\Delta C}$ form I, however, provide us with two new snapshots of the unwinding process. In both of these cases the two *dr*UvrD protomers are located on the same side of the DNA duplex and induce a major bend in the DNA. In *dr*UvrD$^{\Delta C}$, hydrolysis of one of the AMPPNP molecules and release of the nucleotide leads to the formation of crystal form II in which the apo molecule has rotated 125° around the DNA with respect to the position of *dr*UvrD^{FL} or 105° with respect to crystal form I (Figure 29B and C), resulting in an assembly with one *dr*UvrD on either side of the DNA duplex. In this assembly, the apo-*dr*UvrD$^{\Delta C}$ molecule has unwound the 3'-single-stranded DNA extension and maintains the ssDNA in a bent orientation with respect to the DNA duplex axis. As a result, the DNA duplex itself shows a reduced helical twist and reduced bending. Because of the helical nature of nucleic acids, helicases are expected to translocate along DNA in a spiral movement. For the first time, our structures provide three snapshots of this large-scale, ATP-dependent spiral movement. This spiral trajectory results from the positioning of the helicase at an angle relative to the dsDNA axis and from a combination of rotational and translational movements, known as the wrench-and-inchworm mechanism.

Detailed analysis of the ss- and dsDNA binding in our different structures clearly indicates that *dr*UvrD, like *ec*UvrD, utilises the wrench-and-inchworm mechanism for unwinding DNA. In *dr*UvrD$^{\Delta C}$ form I, four nucleotides (nt21-24) are tightly bound in the ssDNA-binding pocket (Figure 30A and B). The terminal nucleotide (nt25) has already exited the helicase and is no longer visible in the electron density maps. Nucleotides 21 and 22 interact with Arg362 and Asn364 and stack against Phe263 that interferes with the regular stacking of the ssDNA bases and forces nucleotides 23 and 24 to adopt an orientation orthogonal to nucleotides 21 and 22. Nucleotide 24 is stabilised in this conformation by π-stacking of the base between Arg264 and Phe196 and of the deoxyribose ring against Phe65. These residues are in turn stabilised by a series of stacking interactions involving notably Tyr261 and His93. In the AMPPNP-bound molecule of *dr*UvrD$^{\Delta C}$ form II, nucleotides 20 to 23 are bound in the binding pocket, indicating that *dr*UvrD has translocated along the ssDNA by one nucleotide compared to form I (Figure 30A and B) and as a result both nucleotides 24 and 25 have become untraceable. As in *ec*UvrD, the apo-*dr*UvrD$^{\Delta C}$ observed in crystal form II, reveals a fifth nucleotide in the ssDNA binding pocket (Figure 30A and B). Nucleotides 19 to 23 are now visible and sliding of *dr*UvrD along the DNA has positioned nucleotides 19 and 20 in the first two binding sites and nucleotides 21 and 22 in the subsequent sites. As a result, nucleotide 23 is now trapped on its way out. To allow the terminal nucleotide to exit, Phe196, Phe65 and His93 have moved out of the way and the α24-α25 linker that interacts via Ser546 with the terminal nucleotide in the AMPPNP–bound forms, has maintained its grip on the 3'-end of the ssDNA and pulled it through the opened gateway driven by the rotation of domain 2B. Nucleotides 21 and 22 are now stabilised in their new binding sites by interactions with Tyr390 and Arg392 from motif IVb.

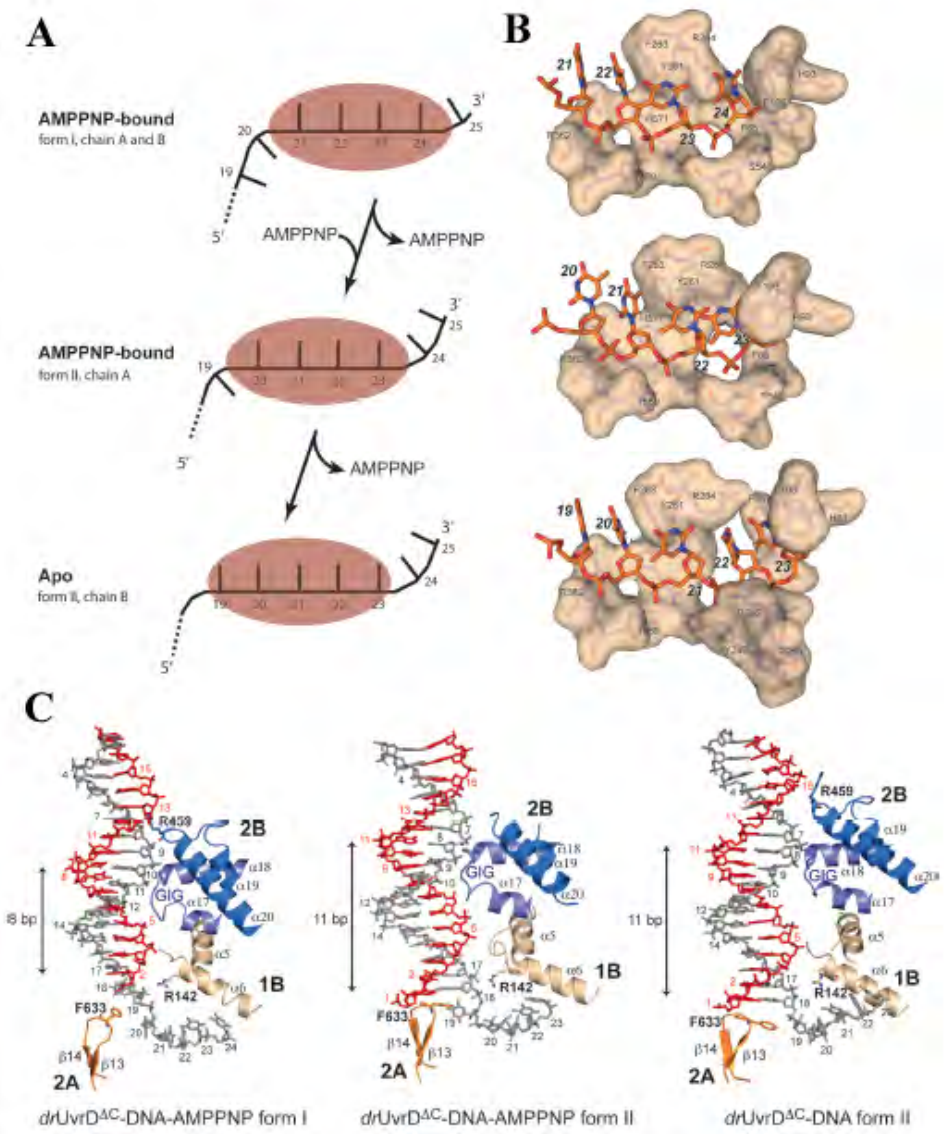


Figure 30: (A) Schematic diagram illustrating the translocation of the different forms of drUvrD$^{\Delta C}$ along the ssDNA. The ssDNA nucleotides are illustrated as black bars and are numbered as in the crystal structures. An oval shape representing drUvrD covers the nucleotides bound in the ssDNA binding pocket. (B) Surface representation of the ssDNA binding pockets of these three forms of drUvrD$^{\Delta C}$ bound to ssDNA (orange sticks). The important residues are labelled and the bases are numbered as in (A). (C) Binding of drUvrD$^{\Delta C}$ to dsDNA in form I, form II with AMPPNP bound and in the apo-form of form II. The dsDNA is illustrated in sticks with the translocated strand in grey. Domains of drUvrD are coloured as in Figure 29A. The helices belonging to the HLH motifs and the β-hairpin structure (orange) are shown and labeled according to the secondary structure succession. The positively charged residues in contact with dsDNA are illustrated in sticks and the GIG motif is indicated. The number of base-pairs formed between the ss-dsDNA junction and the contact point with the drUvrD GIG motif is shown to the left of each panel. This number differs significantly between the two crystal forms.

Similarly the dsDNA binding is affected by the nucleotide-bound state of *dr*UvrD. Interactions between *dr*UvrD and dsDNA involve four contact points: one helix-loop-helix (HLH) motif from domain 1B (α5-α6), two of the three HLH motifs from domain 2B (α17-

α18 and α19-α20) and the β-hairpin motif (β13-β14) from domain 2A (Figure 30C). In the AMPPNP-bound structures, three of these four sites are in contact with dsDNA, two of them are in common and the third differs between the two forms. In both forms, Arg142 from the α5-α6 HLH motif interacts with the unpaired nt19 at the ss-dsDNA junction and the α17-α18 HLH motif containing the conserved GIG motif interacts extensively with nt9-12 in form I and nt7-10 in form II. In form I, the third binding site involves Arg459 from the α19-α20 HLH motif, which interacts with the deoxyribose ring of nt13 (opposite strand) in the minor groove of the DNA duplex, while in form II Phe633 from the β-hairpin motif stacks against the first base-pair (nt1=nt18) of the duplex (Figure 30C). This interaction pushes out nt19 thus forming a bulge in the DNA. In the absence of nucleotide, however, only two of these contacts remain: the GIG sequence in the α17-α18 HLH motif interacts with nt7-10 and Phe633 from the β-hairpin motif stacks against the first base-pair.

Although the details of the protein-DNA contacts are not strictly identical in the structures of *dr*UvrD, *ec*UvrD and *gs*PcrA, taken together, our observations suggest that the molecular mechanisms underlying DNA unwinding are highly conserved within the SF1A helicase superfamily. During ATP-binding-induced domain closing, binding to duplexed DNA (contact 1, see Figure 28A) and ssDNA (contacts 3 and 4) are tight, while contact with the ss-dsDNA junction (contact 2) is loose allowing UvrD to translocate along the duplex. This leads to the unwinding of one nucleotide that forms a bulge at the ss-dsDNA junction. Conversely, during ADP and Pi release, domain rotations open *dr*UvrD's ssDNA gateway (contact 4) to allow the ssDNA to exit the helicase. This is achieved through contacts with the ssDNA (contact 3) that remain tight throughout the process allowing to pull the ssDNA through the gateway and straighten the bulged out nucleotide. During this step, contacts with the DNA duplex are maintained and the β-hairpin (contact 2) stacks against the first base-pair and now acts as a solid separation pin to efficiently unwind the dsDNA.

To better understand how a structurally and mechanistically conserved protein such as UvrD may be involved in diverse repair pathways in different species, we investigated *dr*UvrD's catalytic activities *in vitro*. As other SF1A helicases, *dr*UvrD displays a clear DNA-stimulated ATPase activity. Analysis of the ATPase data measured on *dr*UvrD^{FL} and *dr*UvrD$^{\Delta C}$ allowed us to determine their apparent turnover rates (K_{cat}) for ATP hydrolysis, along with their K_m for ATP and their K_{ssDNA} (corresponding to the concentration of ssDNA required for half-maximal ATPase rate). The values obtained are in agreement with those measured for wild-type and a C-terminally truncated form of *ec*UvrD (Matson and George 1987; Manelyte, Guy et al. 2009).

We then examined the helicase activity of *dr*UvrD on blunt, 3'-tailed, 5'-tailed and both 3'- and 5'-tailed dsDNA (Figure 31). Unexpectedly, our data revealed that *dr*UvrD was most efficient at unwinding dsDNA oligonucleotides containing 5'-extended single-stranded regions. The lowest activities were observed for 3'-extended and blunt dsDNA, while the highest activity was measured with duplexes with extensions at both ends. Such a DNA would provide two entry points for the helicase, which may explain the high helicase activity observed with this substrate. Directionality of SF1 helicases is believed to be determined by preferential binding to either a 3'- or a 5'-single-stranded overhang, which acts as the entry point for the helicase (Singleton, Dillingham et al. 2007). We carried out fluorescence anisotropy measurements to evaluate the affinity of *dr*UvrD for either 3'- or 5'-tailed dsDNA (Figure 31). *dr*UvrD binds to both of these substrates, but shows a 2-fold preference for 5'-extended dsDNA, which may in part explain its higher 5'-3' helicase activity.

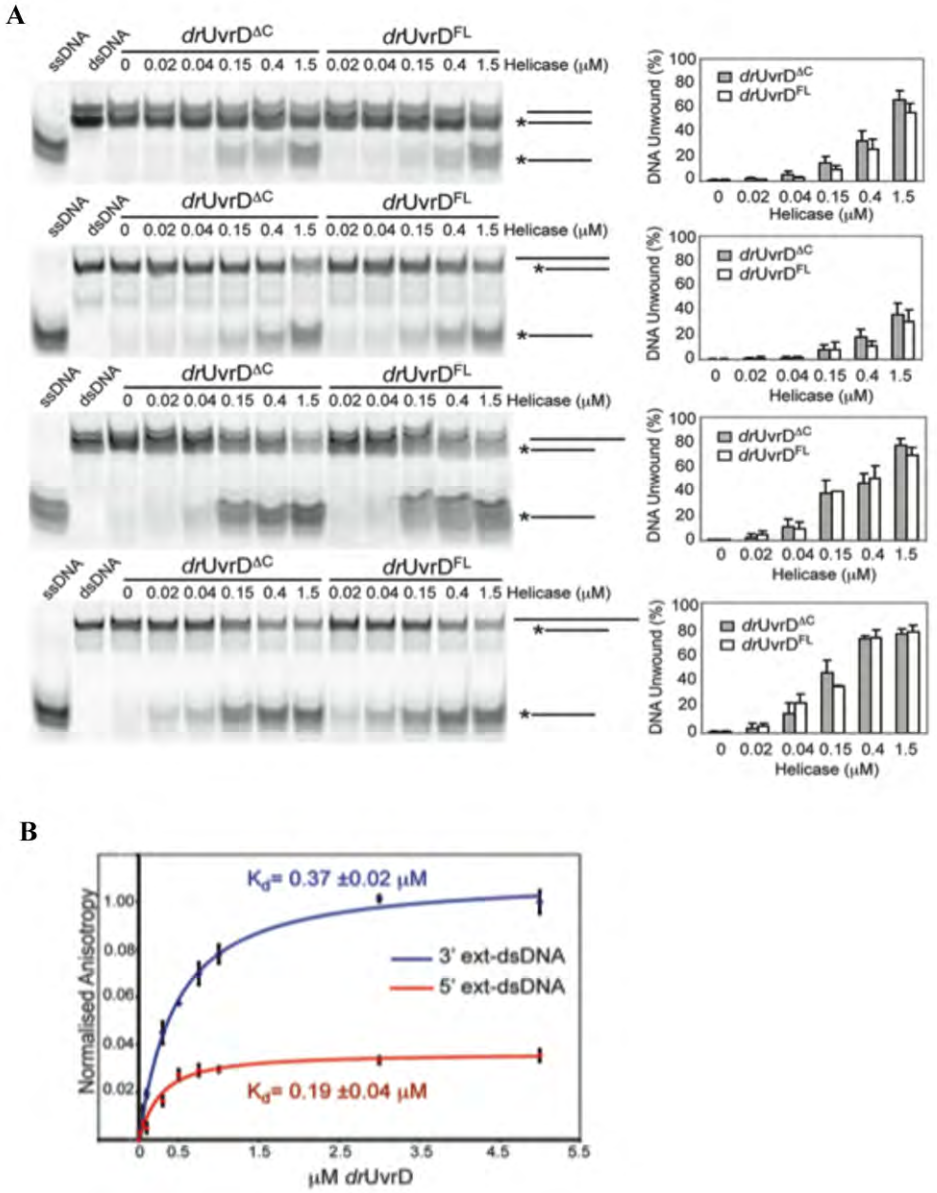


Figure 31: (A) Helicase activity of 0-1.5 μM drUvrD^{FL} and drUvrD$^{\Delta C}$ on 1 μM blunt, 3'-, 5'- and 3'- and 5'-extended dsDNA oligonucleotides. The star indicates the position of the 5'-fluorescein label on the DNA. The gels (left panel) were scanned, the amount of duplexed and single-stranded DNA was quantified and the fraction of unwound DNA was plotted against the amount of helicase (right panel). (B) Fluorescence anisotropy measurements of drUvrD^{FL} binding to 3'- (blue) and 5'- (red) extended dsDNA oligonucleotides. The estimated K_d values for each DNA substrate are shown.

Our and previous structures of UvrD helicases reveal the detailed mechanisms by which they unwind DNA in the 3'-5' direction. There are currently two models that could explain how their directionality may be reversed: (i) in model 1, the contact points between helicase and DNA remain unchanged, but the orientation of the bound DNA is reversed; (ii) in model 2, DNA is bound in the same orientation but the order of alternating tight and loose contacts between helicase and DNA is altered. So far, the high resolution crystal structure of the SF1B helicase, *D. radiodurans* RecD2 (*dr*RecD2) in complex with ssDNA (Saikrishnan, Powell et al. 2009), suggests that 5'-3' translocation is achieved according to the second model. To determine which of these models explains the 5'-3' helicase activity of *dr*UvrD, we

mutated residues identified in *ec*UvrD as being essential for 3'-5' unwinding using the wrench-and-inchworm mechanism and tested the DNA binding and helicase activities of these mutants.

DNA binding was assessed using EMSA (Figure 32). In such an assay a similar fraction of 5'- and 3'-extended DNA was bound by wild-type *dr*UvrD^{FL}. Mutating each of the glycines from the GIG motif individually to threonine (G424T or G426T) did not significantly affect dsDNA binding, however the double mutant (G424T/G426T) showed a near 50% reduction in binding to duplex DNA. No difference was observed between the 5'- and 3'-extended DNA, indicating that the interactions between *dr*UvrD and dsDNA are the same for both DNA substrates. Deletion of the β-hairpin structure, which is known to act as a separation pin and is essential for UvrD's helicase activity, did not interfere with DNA binding. Finally, mutation of Ser546 (S546A) involved in the ssDNA gateway did not affect DNA binding. These mutations had a much more dramatic effect on the helicase activity of *dr*UvrD^{FL} (Figure 32). Deletion of the β-hairpin dramatically reduced DNA unwinding of both 3'- and 5'-extended substrates. The S546A mutant showed clearly reduced activity on 5'-extended DNA only. Mutations in the GIG motif led to reduced activity towards 5'-extended DNA and increased activity towards 3'-extended DNA, resulting in a helicase with reversed polarity. Our mutational study clearly shows that irrespective of the DNA substrate, dsDNA binding occurs through the GIG motif in domain 2B and DNA unwinding requires an intact β-hairpin, indicating that its role as a separation pin is conserved for both 3'- and 5'-extended DNA. *dr*UvrD thus utilises the same wrench-and-inchworm mechanism to unwind DNA in both directions, albeit with higher efficiency in the 5'-3' direction. In the case of *dr*UvrD, reversed polarity is therefore achieved by binding the DNA in the opposite orientation.

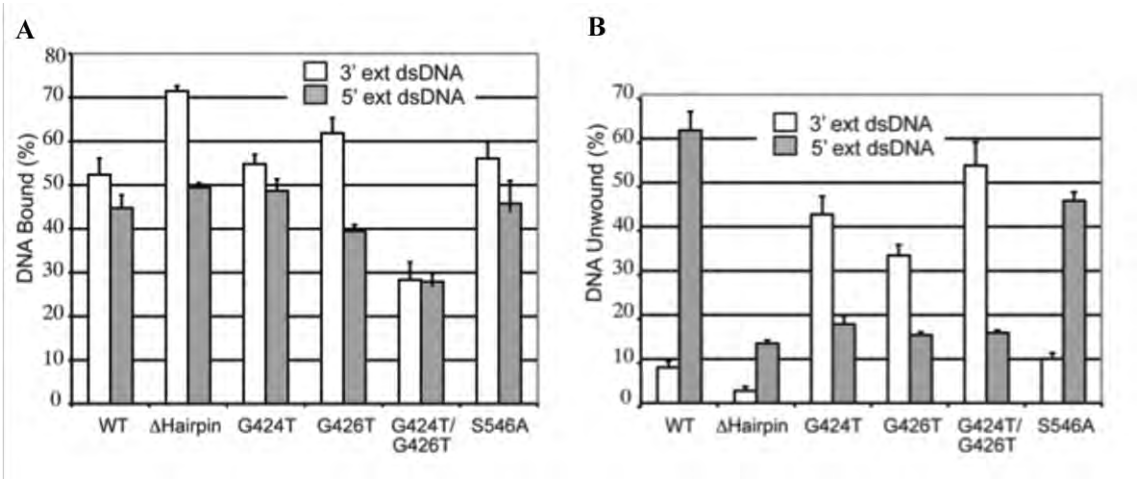


Figure 32: (A) DNA binding activity of 4 μM wild-type (WT) and mutant drUvrD^{FL} on 1 μM 3'- (white) and 5'- (grey) extended dsDNA. (B) Helicase activity of 0.4 μM wild-type (WT) and mutant drUvrD^{FL} on 1 μM 3'- (white) and 5'- (grey) extended dsDNA. The plotted values are the average of at least three independent experiments. Standard deviations are shown as vertical bars.

Despite being structurally and mechanistically conserved with *ec*UvrD and *gs*PcrA, to our surprise, our assays revealed that *dr*UvrD possesses a robust 5'-3' helicase activity and only a weak 3'-5' activity. This is consistent with its preferential binding to 5' extended DNA, which could in part explain its higher 5'-3' helicase activity. Most members of the SF1A family show a clear 3'-5' polarity, but there are, however, several examples, notably in

gram-positive bacteria, of enzymes that show bidirectional helicase activity (Soultanas, Dillingham et al. 1999; Chang, Naqvi et al. 2002; Naqvi, Tinsley et al. 2003; Constantinesco, Forterre et al. 2004; An, Tang et al. 2005). This provides further evidence that UvrD proteins vary both in terms of substrate specificity and helicase polarity. A structural comparison of the ssDNA binding pockets of *ec*- and *dr*UvrD indeed reveals substantial differences, particularly on the side lined by residues from domain 2B, where the conformation and the nature of the residues differ significantly. In *dr*UvrD, the binding pocket forms a positively charged channel that guides the ssDNA towards the exit, while in *ec*UvrD, the pocket is much more open. The residues lining the ssDNA binding pocket and gateway may therefore be determinant in defining the preferred polarity of a given helicase. This region consisting largely of the linker between domains 2B and 2A (residues 540-550 in *dr*UvrD) is poorly conserved in SF1A helicases and could account for the observed variations in substrate specificity. This hypothesis is supported by our S546A mutant, which displays reduced activity on 5'-extended DNA and not on 3'-tailed DNA.

*dr*UvrD's ability to unwind DNA duplexes in both directions may reflect its implication in diverse DNA repair pathways *in vivo*. In *E. coli*, recombinational repair has been proposed to involve the 3'-5' helicases RecQ and Helicase IV and the 5'-3' helicase RecD, while *D. radiodurans* cells missing these genes show wild-type radioresistance and DNA repair capacity (Cao and Julin 2009; Bentchikou, Servant et al. 2010). In contrast, inactivation of *dr*UvrD leads to a significant increase in the sensitivity of cells to γ-irradiation. This phenotype is further enhanced in cells in which both *uvrD* and *recD2* genes have been disrupted, suggesting that the 5'-3' helicase, *dr*RecD2, may in part back-up *dr*UvrD's function. While further studies will be needed to decipher the detailed molecular mechanisms that regulate the bipolar helicase activity of *dr*UvrD, these observations suggest that *in vivo* both helicase activities of *dr*UvrD are needed. *dr*UvrD may be able to switch from a highly coordinated 5'-3' helicase to a processive 3'-5' helicase in response to external stresses, changes in its environment, or upon interactions with pathway-specific protein partners such as MutL(Matson and Robertson 2006)(Matson and Robertson 2006) or UvrAB.

3. Base excision repair

Since 2003, the ESRF Structural Biology Group has been collaborating with the Norwegian Structural Biology Centre (NorStruct), a national research and service centre located in Tromsø (Norway). Two scientists from NorStruct, Ingar Leiros and Elin Moe, led by Arne Smalås, have been studying the Base Excision Repair (BER) enzymes of several bacterial species including those of *D. radiodurans*. I. Leiros was a post-doctoral fellow at ESRF from 2002 to 2005, while E. Moe was a regular visitor of the ESRF laboratory between 2003 and 2010. During this time, E. Moe cloned the 13 predicted DNA glycosylases encoded by the *D. radiodurans* genome and managed to express and purify a large majority of them (Table 9). So far, together with her co-workers, E. Moe has solved the structures of five of these, three of which have been published (Leiros, Moe et al. 2005; Moe, Leiros et al. 2006; Moe, Hall et al. 2012).

From 2006 onwards, I became more involved in E. Moe's projects and was her principal contact at the ESRF. In 2007, E. Moe and I submitted a grant proposal to the Norwegian Research council, entitled: 'Functional and structural biology studies of DNA repair proteins from the radiation resistant bacterium Deinococcus radiodurans'. This project was successful and we obtained funding for a PhD student (Aili Sarre), who started in 2008 under the supervision of E. Moe in Tromsø and myself during her regular visits to Grenoble.

***Table 9**: Status of the cloning, expression, purification, crystallisation and structure determination of D. radiodurans DNA glycosylases.*

Protein	Code	Cloned	Expressed	Purified	Crystallised	Structure determined
UNG	**DR0689**	**X**	**X**	**X**	**X**	**X**
MUG	**DR0715**	**X**	**X**	**X**	**X**	**X**
TMUDG	**DR1751**	**X**	**X**	**(X)**		
Hypothetical UDGs	**DR0022** **DR1663**	**X** **X**	**X** **X**	**X** **X**		
AlkA	**DR2074** **DR2584**	**X** **X**	**X** **X**	**(X)** **X**	**X**	**X**
EndoIII	**DR2438** **DR0289** **DR0928**	**X** **X** **X**	**X** **X** **X**	**X** **X** **X**	**X** **X**	**X** **X**
MutM	**DR0493**	**X**	**X**	**X**		
MutY	**DR2285**	**X**	**(X)**			
MutT	**DR0261**	**X**	**X**	**(X)**		

A. Sarre's project focuses on the study of *D. radiodurans'* Endonuclease III enzymes. *D. radiodurans* possesses an expanded repertoire of BER enzymes: for example its genome encodes two 3-methyladenine DNA glycosylases (AlkA) and three Endonuclease III enzymes (EndoIII), while most bacteria only possess one copy of each of these essential BER enzymes.

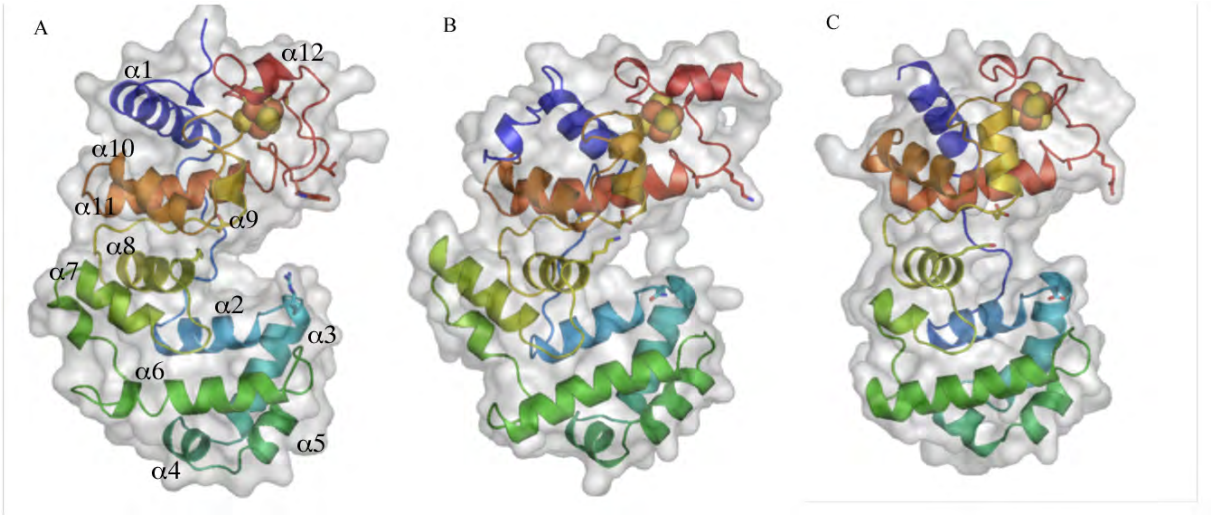


Figure 33: (A)-(B) Ribbon illustrations of the crystal structures of D. radiodurans Endonuclease III-1 (A) and ΔN76-Endonuclease III-3 (B). (C) Homology model of D. radiodurans Endonuclease III-2. The structures are coloured in rainbow with the N-terminus being blue and the C-terminus red. The helices are numbered in the case of EndoIII-1. Active site and DNA binding residues are shown in sticks.

The BER pathway is responsible for repair of DNA bases damaged by deamination, oxidation and alkylation. The pathway is initiated by DNA glycosylases, which recognise the

damaged bases and remove them by hydrolysing the N-glycosidic bond between the base and the sugar phosphate backbone and generates an abasic site. The after base removal process are performed by apurinic (AP) endonucleases, DNA polymerases and DNA ligases. Some of the DNA glycosylases are bi-functional and contain an AP-lyase activity in addition to their N-glycosidic activity, and are not dependent upon the lyase activity of DNA polymerase in order to complete the repair.

Table 10: *Data and refinement statistics of D. radiodurans EndoIII-1 and ΔN76-EndoIII-3 structures. Values in parentheses are for the highest resolution shell.*

Data collection	**EndoIII-1**	**ΔN76-EndoIII-3**
Space group	C2	C2
Cell dimensions (Å/º)	a=181.50, b=38.56, c=37.11, β=89.4	a=91.24, b=40.21, c=72.15, β=102.3
Resolution (Å)	19.5-1.95 (2.00-1.95)	50-1.31 (1.34-1.31)
R_{merge} (%)	5.8 (28.7)	5.8 (30.2)
Unique reflections	18638	61009
Completeness (%)	98.7 (88.7)	98.2% (90.2%)
Nº mol/asu	1	1
Refinement		
R_{factor}/ R_{free} (%)	17.5/23.3	13.5/16.7
No. atoms	2016	2313
R.m.s. deviations		
Bond lengths (Å)	0.015	0.023
Bond angles (º)	1.301	1.985

EndoIII is a bifunctional enzyme that recognises oxidised pyrimidines and belongs to the HhH-GPD superfamily of DNA-glycosylases. The first EndoIII protein to be both functionally and structurally characterised was *E. coli* EndoIII. It consists of two domains with four N-terminal α-helices and a C-terminal six α-helix barrel. Between the two domains is a solvent filled groove containing the active site and in which the DNA is bound (Fromme and Verdine 2003). Both domains have well-conserved DNA interacting motifs, a Helix-hairpin-Helix motif (HhH) and a [Fe_4S_4] cluster loop, which is unique to the HhH-GDP family proteins. The metal cluster is a redox-inactive cubane [4Fe-4S]-cluster coordinated in position by a loop with four conserved cysteines (C-X_6-C-X_2-C-X_5-C).

A. Sarre is now in the process of writing-up her PhD thesis in the form of three manuscripts, in which she describes the crystal structures of two of the three EndoIIIs of *D. radiodurans* (EndoIII-1 and EndoIII-3) and the *in vitro* characterisation of the catalytic activity and substrate specificity of these three enzymes. Despite numerous tries, crystals of EndoIII-2 could not be obtained. Instead, T. Klar (post-doc) and A. Sarre respectively solved

the structures of intact EndoIII-1 and of an N-terminally truncated EndoIII-3 (ΔN76-EndoIII-3), with the help of M. Ökvist (Figure 33). In both cases, the structures were solved by SAD method making use of the bound iron as an intrinsic anomalous scatterer. The EndoIII-1 and ΔN76-EndoIII-3 structures were refined respectively to 1.95 and 1.31 Å resolution (Table 10). In the absence of a crystal structure for EndoIII-2, a homology model was prepared using *E. coli* EndoIII as a basis for the modelling, which shares 47% sequence identity with *D. radiodurans* EndoIII-2 (Figure 33). Overall, the structures of *E. coli* and *D. radiodurans* EndoIIIs are very similar. However, there are a number of differences particularly in the DNA binding groove, where a number of active site residues and DNA binding residues differ from one enzyme to the other.

To better understand the importance of these small structural differences, A. Sarre has investigated the ability of the three *D. radiodurans* EndoIII enzymes to recognise and remove thymine glycol (Tg; a known substrate of EndoIII) from DNA. Interestingly, the three EndoIIIs are very distinct in terms of activity. Based on sequence alignments, EndoIII-2 is the most similar to *E. coli* and human EndoIII, and indeed it is also an active bifunctional enzyme on Tg-containing DNA (Figure 34). In contrast, EndoIII-1 only possesses DNA glycosylase activity on Tg-DNA and not AP-lyase activity as expected for a bifunctional glycosylase. EndoIII-1 is therefore only a monofunctional enzyme. Finally EndoIII-3 appears to have no activity on Tg-DNA. We now hope to identify additional and perhaps novel substrates for these three enzymes.

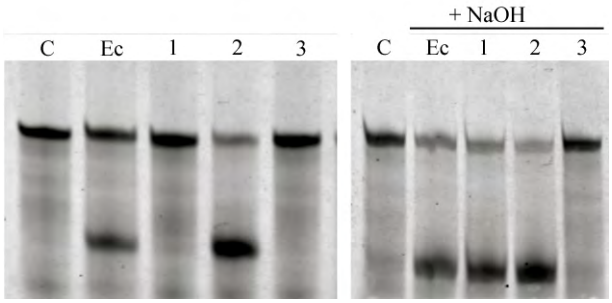


Figure 34: Cleavage of Thymine glycol (Tg) containing DNA. A 35-mer duplex oligonucleotide containing a single Tg at position 14 and 5' labelled with 6-FAM was incubated for 1 hour at 37ºC with either no enzyme (C), E. coli EndoIII (Ec), D. radiodurans EndoIII-1 (1), EndoIII-2 (2) or EndoIII-3 (3). Cleaved DNA was separated from intact DNA by 20% denaturing urea-PAGE and visualised on a fluorescence imager. On the right, prior to electrophoresis, NaOH was added to each of the samples containing enzyme, which causes abasic site containing DNA to be cleaved. This allows differentiating between mono- and bi-functional activities of EndoIIIs.

Since 2010, I have been collaborating with Jean-François Constant and Muriel Jourdan from the University Joseph Fourier in Grenoble on the study of MutM (also known as Fpg, Formamidopyrimidine glycosylase) from *D. radiodurans*. A PhD student, Morgane Lourdin, has expressed and purified the protein and the aim of her project is to carry out structural (NMR and X-ray crystallography) and functional studies on this enzyme. The establishment of a reliable purification protocol was carried out at ESRF and since 2011 at IBS under my supervision. We are now trying to obtain crystals of the enzyme alone and in complex with abasic site containing DNA oligonucleotides.

Part III: Research Project

The prime objective for every life form is to deliver its genetic material, intact and unchanged, to the next generation, despite constant assaults from both endogenous and environmental sources on the DNA. The average human cell suffers over 10,000 DNA lesions per day (Ames, Shigenaga et al. 1993). If left unrepaired, damaged DNA generates mutations, replication errors, persistent DNA damage and genomic instability that ultimately threaten cell or organism viability. To counter this threat, cells have evolved several elaborate DNA damage response systems. The wide diversity of DNA lesions necessitates multiple, largely distinct DNA repair mechanisms (Jackson and Bartek 2009). Deletions/insertions and specific base modifications (oxidation, methylation) are repaired via the mismatch repair and BER pathways respectively, while UV-mediated damage is repaired via the NER pathway. In the case of double-strand break repair, two mechanisms are known to be involved, namely non-homologous end-joining and homologous recombination. The principles and underlying mechanisms of DNA repair are largely conserved throughout evolution, though the complexity of the systems varies immensely. This is largely due to the requirement for a coordinated response to DNA damage in higher eukaryotes - a global mechanism known as DNA Damage Response (DDR) – a process that includes damage recognition, intracellular signalling via protein kinases, DNA damage repair and regulation of diverse cellular responses such as the cell cycle, chromatin remodelling and gene transcription (Jackson and Bartek 2009).

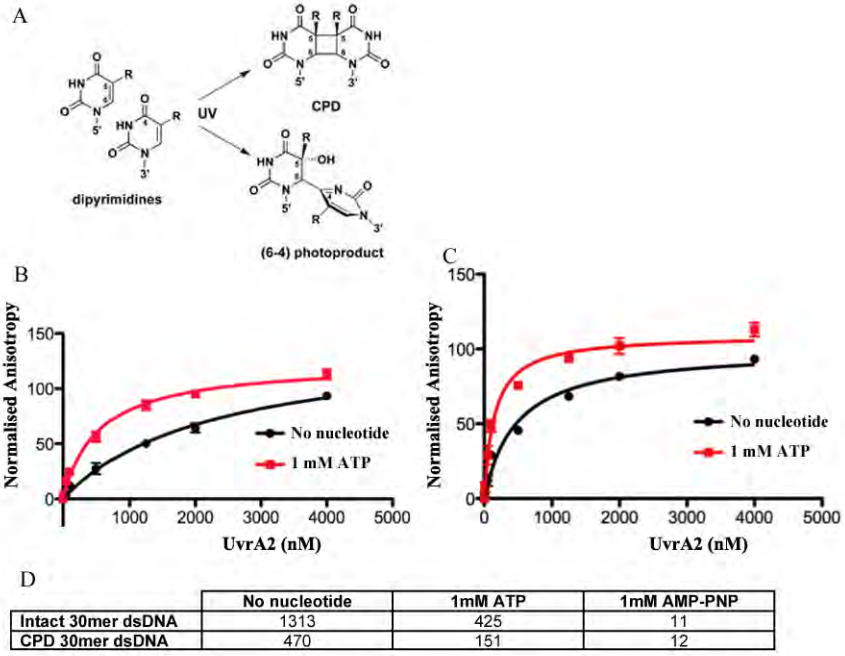
Mutations of the genes involved in DNA repair are directly responsible for several human syndromes and are associated with a dramatic increase in predisposition to cancer. A hallmark of cancer is genome instability and most carcinogens operate by generating DNA damage and causing mutations (Jackson and Bartek 2009). Paradoxically, the ability of cancer cells to repair DNA damage after chemo- or radiotherapy significantly reduces the efficiency of available treatments. As a result, proteins involved in DNA repair pathways are now being viewed as potential drug targets. More recently, a number of viruses have been shown to interfere with host cellular DNA repair pathways, by either trying to evade the cellular response or by hijacking the cellular damage response pathway to facilitate their own replication. These observations suggest that the interactions of these viral proteins with the host DNA repair machinery constitute an attractive target for drug development. In the recent years, small molecule inhibitors of DNA repair proteins have successfully been used to potentiate the effect of radiotherapy in the treatment of several cancers and to significantly reduce the ability of HIV-1 to replicate in cells. At present, screening and designing specific inhibitors for these new drug targets are largely dependent on the availability of high-resolution crystal structures of these essential DNA repair enzymes and a more detailed understanding of the molecular mechanisms underlying these processes.

The primary interest of my newly established team at IBS is the study of the molecular mechanisms underlying DNA damage recognition and repair in humans and in the radiation-resistant bacterium *Deinococcus radiodurans*. Our work focuses on two major aspects, which are ***(i)*** **Recognition of DNA Lesions** and ***(ii)*** **Dynamics of DNA Damage Repair Processes**. The main objectives will be to determine the crystal structures of several DNA repair enzymes alone or in complex with their DNA substrates or protein partners and to study the dynamics of DNA damage recognition and protein assembly at sites of DNA lesion. In the long term, these structures will be analysed in light of known mutations in these genes associated with cancer predispositions and will be used to screen and design small molecule inhibitors that could be used to either reduce viral replication or potentiate the efficacy of current anti-cancer drugs. The goal is to use a combination of Structural Biology methods,

Biophysical and Biochemical tools and single-molecule fluorescence techniques to decipher the complex molecular processes leading to efficient repair of DNA lesions.

A. Recognition of DNA Lesions

DNA damage recognition is a very challenging cellular process, which consists in detecting very small modifications to DNA bases in a large pool of intact DNA. Ultra-violet (UV) light is the most pervasive environmental DNA-damaging agent and introduces a diversity of lesions. NER is the primary pathway for repair of such lesions and in prokaryotes, recent studies reveal that efficient recognition of such a broad set of lesions is largely accomplished by a single protein, UvrA. In contrast the BER enzymes, responsible for repairing lesions associated with oxidative stress, are very different in that they are multiple (13 DNA glycosylases in *D. radiodurans*) and each one of them recognises only a very small set of chemically related DNA lesions and hydrolyse the N-glycosidic bonds of the aberrant base. We would therefore like to address the following questions: how does a small set of NER proteins recognise a broad range of structurally diverse DNA lesions or DNA distortions? How do the specialised BER enzymes recognise small chemical modifications to bases within a large pool of undamaged DNA?



	No nucleotide	1mM ATP	1mM AMP-PNP
Intact 30mer dsDNA	1313	425	11
CPD 30mer dsDNA	470	151	12

Figure 35: (A) Structures of the two principal DNA lesions induced by UV light. CPD: cyclobutane pyrimidine dimer and (6-4) photoproduct. (B)-(C) Fluorescence anisotropy measurements of UvrA2 binding to intact (B) and CPD-containing (C) 30mer dsDNA oligonucleotides in the absence (black line) and presence of 1 mM ATP (red line). Similar measurements were carried out in the presence of 1 mM AMPPNP (not shown). (D) Dissociation constants (Kd, nM) derived from the data presented in (B) and (C).

To answer these questions, our initial goal will be to study the substrate specificity and determine the high-resolution crystal structures of *D. radiodurans* UvrAs and two human BER enzymes (hNTH1 and hMYH), associated with increased risks of cancer, in complex with oligonucleotides containing DNA lesions. This project will greatly benefit from our newly established collaboration with Jean-Luc Ravanat and Didier Gasparutto from the 'Lésions des Acides Nucléiques' laboratory of the CEA in Grenoble, who are world-leaders in

the synthesis of lesion-containing oligonucleotides and the identification and characterisation of damaged DNA bases. In the long-term we will also analyse these crystal structures in view of known mutations in these genes associated with cancer predispositions and use them to design specific inhibitors that could be used to potentiate the efficacy of drugs currently used in chemo- or radiotherapy.

In the case of the NER pathway, we will initially focus on determining the substrate specificity of *D. radiodurans* UvrAs (UvrA1 and UvrA2) and the role of ATP binding and hydrolysis in the processes of DNA binding and damage recognition. Identifying good DNA substrates for these proteins will allow us to initiate crystallisation trials to obtain high resolution crystal structures of UvrAs in complex with damaged DNA in order to map the specific interactions required for damage recognition and better understand the molecular mechanisms involved in regulating this ATP-dependent process. Figure 35 illustrates our first results on the study of UvrA2 binding to intact and cyclobutane pyrimidine dimer (CPD)-containing 30mer dsDNA oligonucleotides. The presence of a nucleotide (ATP or AMPPNP) increases the affinity of UvrA2 for dsDNA. UvrA2 binds with higher affinity to damaged DNA in the absence of nucleotide or with ATP, while the addition of non-hydrolysable ATP (AMPPNP) locks the protein on the DNA and blocks its ability to discriminate between intact and damaged DNA.

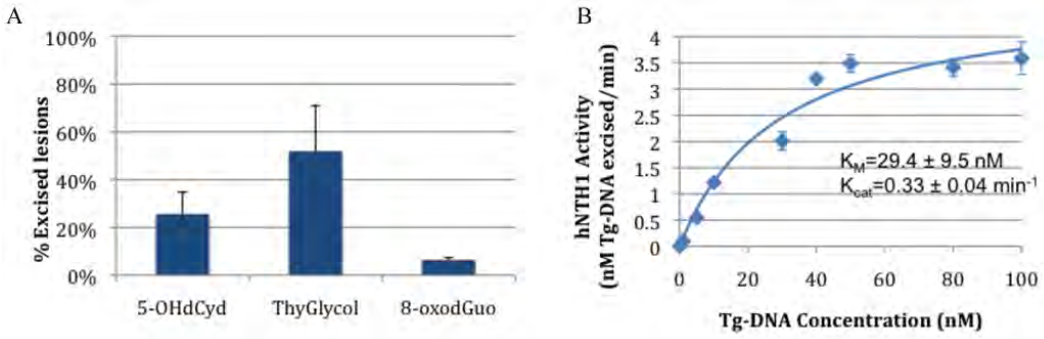


Figure 36: (A) hNTH1 excision activity on γ-irradiated (75 Gy) genomic DNA. The histogram shows the fraction of 5-hydroxycytosine (5-OHdCyd), Thymine glycol (ThyGlycol) and 8-oxoguanine (8-oxodGuo) excised by hNTH1 in 1 hour at 37°C. (B). Enzymatic activity of 15nM hNTH1 on 0 to 100nM 35 base-pair (bp) oligonucleotide containing a thymine glycol lesion in position 14. Our data points were fitted to a standard Michaelis-Menten curve and estimated K_M and K_{cat} values are shown.

In the case of the BER pathway, our initial work has focused on the human Endonuclease III enzyme (hNTH1). A codon-optimised full-length construct of hNTH1 was expressed in *E. coli* and a purification protocol was established yielding milligrams of active enzyme. A Masters student, S. Gauthier, characterised hNTH1's enzymatic activity on thymine glycol (Tg) containing dsDNA oligonucleotides in order to study the kinetics of the reaction (Figure 36). In collaboration with J.-L. Ravanat from the CEA, we have also confirmed that hNTH1 can specifically excise Tg and 5-hydroxycytosine from irradiated genomic DNA using liquid chromatography coupled to mass spectrometry analysis (Figure 36). New DNA substrates are currently being synthesised in order to further investigate the substrate specificity of hNTH1 and identify suitable substrates for co-crystallisation. In the case of bifunctional DNA glycosylases, stable protein-DNA complexes can be obtained either by introducing mutations into the protein to make it inactive or by making use of their ability to form an irreversible covalent attachment to a DNA substrate in the presence of borohydride ion through the reduction of the Schiff base intermediate.

B. Dynamics of DNA Damage Repair Processes

In addition to our high-resolution structural work, we wish to investigate the dynamics of DNA repair processes and the critical protein-protein and protein-DNA interactions that regulate DNA damage recognition and repair in order to define more precisely the sequence of events leading from lesion detection to efficient repair.

We will use single-molecule fluorescence techniques to study the dynamics of NER by the Uvr proteins and more particularly the early steps of NER involving damage recognition by UvrA and UvrB proteins. In collaboration with Dr. Virgile Adam from the IBS in Grenoble and Robert Neely from the University of Leuven in Belgium, we are developing a fluorescence-based *in vitro* assay in which we will study the interactions (using Förster resonance energy transfer, FRET) and dynamics (single-molecule imaging) of this process. A few months ago we also initiated experiments to look at the dynamics of the Uvr repair system *in vivo* using super-resolution microscopy (Photoactivated localisation microscopy, PALM). A Masters student, Rémy Flores-Flores supervised by V. Adam carried out preliminary experiments in *E. coli* to identify the best photoconvertible fluorescent proteins to fuse to our proteins of interest. We obtained very nice images using either Dendra2 or mEOSFP-M159A (Adam, Moeyaert et al. 2011) as fluorescent proteins (Figure 37). The localisation is very reminiscent of nucleoid staining previously seen in super-resolution microscopy images of the nucleoid-associated protein HU (Lee, Thompson et al. 2011) (Figure 37C). Experiments are now underway to introduce genes encoding these fluorescent proteins downstream of the two UvrA genes and the UvrB gene in *D. radiodurans'* genome in order to follow their localisation in response to UV irradiation. The genetic modification of *D. radiodurans* is carried out in collaboration with Prof. Suzanne Sommer from Orsay.

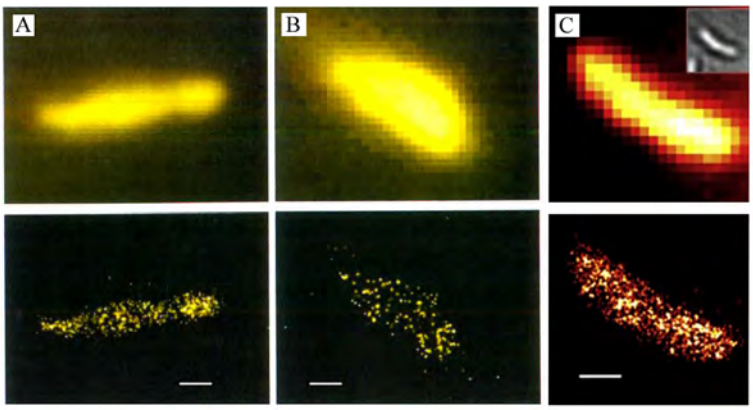


Figure 37: (A)-(B) Diffraction-limited images (top) and super-resolution images (bottom) of Dendra2-UvrA1 (A) and mEOSFP-M159A-UvrA1 (B) in fixed E. coli cells overexpressing UvrA1 in fusion with either Dendra2 or mEOSFP-M159A (Adam, Moeyaert et al. 2011). (C) Example of diffraction-limited (top) and super-resolution (bottom) images of the nucleoid-associated protein HU fused to eYFP in fixed C. crescentus (Lee, Thompson et al. 2011). Scale bars represent 0.5µm.

In 2011, the successful expression and purification of *D. radiodurans* UvrC allowed us to initiate interaction studies with UvrB. Unlike UvrA and UvrB, which do not form a stable complex in our hands, we were able to isolate a stable UvrB-UvrC complex and start to characterise such a complex (Figure 38). The complex seems to form in a 1:1 stoichiometry and, unlike the individual proteins, binds very efficiently to fluorescein-containing dsDNA.

Initial crystallisation trials of a ternary complex containing UvrB, UvrC and dsDNA were carried out and yielded some promising hits, which now need to be improved.

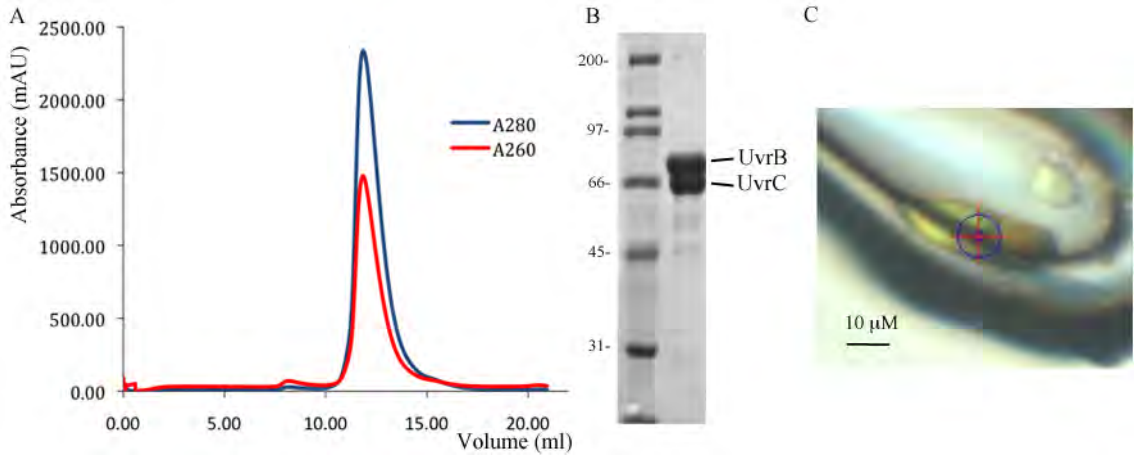


Figure 38: (A) SEC profile of the UvrB-UvrC complex from D. radiodurans, separated on a Superdex 200 column. (B) SDS-PAGE analysis of the UvrB-UvrC complex. (C) Initial crystal hits of UvrB-UvrC-dsDNA obtained from the crystallisation robot. The crystal is mounted in a cryoloop and centered in the beam on the microfocus beamline at the ESRF.

In humans, the Y-box-binding protein 1 (YB-1), which is a multifunctional transcription factor, has recently been shown to interact directly with hNTH1 and stimulate its DNA repair activity. YB-1 is directly involved in the cellular response to genotoxic stress and is up-regulated in tumour cell lines resistant to cisplatin (Bargou, Jurchott et al. 1997; Janz, Harbeck et al. 2002). The abundance of YB-1/hNTH1 complex was shown to increase in tumour cells treated with UV light and cisplatin and siRNA specific to hNTH1 mRNA increases cisplatin-sensitivity in YB-1 overexpressing cells (Guay, Garand et al. 2008). These findings indicate that the enzymatic activity of hNTH1 constitutes a potential drug target against cisplatin-resistant and perhaps also radiation-resistant tumours. In order to improve our understanding of the role of YB-1 and hNTH1 in the cellular response to genotoxic stress, we also wish to further characterise the YB-1/hNTH1 complex at both a biochemical (*in vitro*) and cellular (*in vivo*) level. We will carry out the biophysical characterisation of the complex using multi-angle laser light scattering coupled to SEC to estimate its size and stoichiometry, and Surface Plasmon Resonance or Isothermal Titration Calorimetry to estimate binding constants. We will also study the complex by SAXS to obtain low-resolution envelopes and will submit the purified YB-1/hNTH1 complex (or complexes of relevant protein domains) to the local crystallisation platform in Grenoble in order to obtain crystals suitable for diffraction experiments.

Finally, we would like to further characterise the interactions and dynamics involved in the formation of the RecO-RecR and the RecR-RecF complexes from *D. radiodurans* and try to understand the role of DNA in this process. Since we have already determined two crystal structures of the RecO-RecR complex, we will focus on obtaining crystals of the RecR-RecF complex and will use a combination of SANS and SAXS to characterise the different assemblies formed in the presence of DNA. Of course, this structural work will be supported by biochemical and biophysical studies of these complexes *in vitro* so as to determine the stoichiometry and oligomeric states of the various subunits in the complexes. Our work on *D. radiodurans* RecN will also be continued with a focus on the characterisation of the full-length RecN. Our work will focus on (i) identifying the areas required for DNA

binding and DSB recognition, (ii) characterising its oligomeric state and ability to polymerise along DNA and (iii) identifying cellular binding partners of RecN.

The long-term perspectives of this project are:

1. Improving our understanding of the detailed mechanisms underlying DNA repair and recognition of DNA lesions – a very challenging process which involves finding small modifications to bases in a sea of undamaged DNA.

2. Identification of new anti-cancer drug targets and design of specific inhibitors based on high-resolution crystal structures of such target enzymes.

This research project is currently supported by an ATIP-AVENIR grant from the CNRS for 3 years (2011-2014) and is additionally sponsored by La Ligue Contre le Cancer. Funding from the Département du Science de Vivant of the CEA was obtained in 2011 to support our work on the functional characterisation of the Endonuclease III enzymes and a grant from the Association pour la Recherche sur le Cancer (ARC) was obtained in 2012 for our structural studies of DNA repair proteins in complex with lesion-containing DNA.

References:

Adam, V., B. Moeyaert, et al. (2011). Chem Biol 18(10): 1241-1251.
Ali, J. A. and T. M. Lohman (1997). Science 275(5298): 377-380.
Altenberg, G. A. (2003). News in Physiological Sciences 18: 191-195.
Ames, B. N., M. K. Shigenaga, et al. (1993). Proc Natl Acad Sci U S A. 90(17): 7915-7922.
An, L., W. Tang, et al. (2005). J Biol Chem 280(32): 28952-28958.
Andersen, C. B., T. Becker, et al. (2006). Nature. 443(7112): 663-668.
Aravind, L., D. D. Leipe, et al. (1998). Nucleic Acids Res 26(18): 4205-4213.
Arthur, H. M. and R. G. Lloyd (1980). Mol Gen Genet 180(1): 185-191.
Bargou, R. C., K. Jurchott, et al. (1997). Nat Med 3(4): 447-450.
Bentchikou, E., P. Servant, et al. (2010). PLoS Genet 6(1): e1000774.
Bochkarev, A. and E. Bochkareva (2004). Curr Opin Struct Biol 14: 36-42.
Branden, C. and J. Tooze (1991). Introduction to Protein Structure, Garland Publishing, Inc. NY.
Bruand, C. and S. D. Ehrlich (2000). Mol Microbiol 35(1): 204-210.
Cao, Z. and D. A. Julin (2009). DNA Repair (Amst) 8(5): 612-619.
Caron, P. R., S. R. Kushner, et al. (1985). Proc Natl Acad Sci U S A. 82(15): 4925-4929.
Chang, T. L., A. Naqvi, et al. (2002). J Biol Chem 277(48): 45880-45886.
Cheng, W., K. M. Brendza, et al. (2002). Proc Natl Acad Sci U S A 99(25): 16006-16011.
Chow, K. H. and J. Courcelle (2004). J Biol Chem 279(5): 3492-3496.
Connelly, J. C. and D. R. Leach (2002). Trends Biochem Sci 27(8): 410-418.
Constantinesco, F., P. Forterre, et al. (2004). Nucleic Acids Res 32(4): 1439-1447.
Cox, M. M. and J. R. Battista (2005). Nat Rev Microbiol 3(11): 882-892.
Cromie, G. A. and D. R. Leach (2001). Mol Microbiol 41(4): 873-883.
Croteau, D. L., M. J. DellaVecchia, et al. (2008). DNA Repair (Amst) 7(3): 392-404.
Croteau, D. L., M. J. DellaVecchia, et al. (2006). J Biol Chem 281(36): 26370-26381.
Cuypers, M. G., E. P. Mitchell, et al. (2007). J Mol Biol 371(3): 787-799.
Daly, M. J., E. K. Gaidamakova, et al. (2007). PLoS Biol 5(4): e92.
DellaVecchia, M. J., D. L. Croteau, et al. (2004). J Biol Chem. 279(43): 45245-45256.
Dennis, R. J., E. Micossi, et al. (2006). Acta Crystallogr Sect F Struct Biol Cryst Commun 62(Pt 4): 325-329.
Dessinges, M. N., T. Lionnet, et al. (2004). Proc Natl Acad Sci U S A 101(17): 6439-6444.
Dillingham, M. S., D. B. Wigley, et al. (2002). Biochemistry 41(2): 643-651.
Elbein, A. D., Y. T. Pan, et al. (2003). Glycobiology 13(4): 17R-27R.
Elser, V. (2003). Journal of the Optical Society of America a-Optics Image Science and Vision 20(1): 40-55.
Fischer, C. J., N. K. Maluf, et al. (2004). J Mol Biol 344(5): 1287-1309.
Fromme, J. C. and G. L. Verdine (2003). EMBO J 22(13): 3461-3471.
Funayama, T., I. Narumi, et al. (1999). Mutation Research-DNA Repair 435(2): 151-161.
Gomis-Roth, F. X., A. Dessen, et al. (2003). Structure 11(4): 423-433.
Goosen, N. and G. F. Moolenaar (2008). DNA Repair (Amst) 7(3): 353-379.
Gorbalenya, A. E. and E. V. Koonin (1993). Current Opinion in Structural Biology 3(3): 419-429.
Graumann, P. L. (2001). Biochimie 83(1): 53-59.
Graumann, P. L. and T. Knust (2009). Chromosome Res 17(2): 265-275.
Grove, J. I., S. R. Wood, et al. (2009). DNA Repair 8(12): 1434-1443.
Guay, D., C. Garand, et al. (2008). Cancer Sci 99(4): 762-769.
Hegde, S. P., M. H. Qin, et al. (1996). Proc Natl Acad Sci USA 93(25): 14468-14473.
Higgins, C. F. (1992). Annu Rev Cell Biol. 8: 67-113.

Hirano, T. (2002). Genes & Development 16(4): 399-414.
Hirano, T. (2005). Philosophical Transactions of the Royal Society B-Biological Sciences 360(1455): 507-514.
Hirano, T. (2006). At the heart of the chromosome: SMC proteins in action. Nature Reviews Molecular Cell Biology. 7: 311-322.
Hoenen, T., V. Volchkov, et al. (2005). Journal of Virology 79(3): 1898-1905.
Honda, M., J. Inoue, et al. (2006). J Biol Chem. 281(27): 18549-18559.
Hopfner, K. P., L. Craig, et al. (2002). Nature 418(6897): 562-566.
Hopfner, K. P., A. Karcher, et al. (2000). Cell. 101(7): 789-800.
Hopfner, K. P. and J. A. Tainer (2003). Curr Opin Struct Biol. 13(2): 249-255.
Jackson, S. P. and J. Bartek (2009). Nature 461(7267): 1071-1078.
Janz, M., N. Harbeck, et al. (2002). Int J Cancer 97(3): 278-282.
Kanaar, R., J. H. J. Hoeijmakers, et al. (1998). Trends in Cell Biology 8(12): 483-489.
Kapp, U., S. Macedo, et al. (2008). Acta Crystallogr Sect F Struct Biol Cryst Commun 64(Pt 6): 479-486.
Karakas, E., J. J. Truglio, et al. (2007). EMBO J 26(2): 613-622.
Karpowich, N., O. Martsinkevich, et al. (2001). Structure. 9(7): 571-586.
Kidane, D., H. Sanchez, et al. (2004). Molecular Microbiology 52(6): 1627-1639.
Korolev, S., J. Hsieh, et al. (1997). Cell 90(4): 635-647.
Kosa, J. L., Z. Z. Zdraveski, et al. (2004). Mutation Research-Fundamental and Molecular Mechanisms of Mutagenesis 554(1-2): 149-157.
Kowalczykowski, S. C., D. A. Dixon, et al. (1994). Microbiol Rev 58(3): 401-465.
Kuzminov, A. (1999). Microbiol Mol Biol Rev 63(4): 751-813, table of contents.
Lamarche, B. J., N. I. Orazio, et al. (2010). FEBS Lett 584(17): 3682-3695.
Lamers, M. H., A. Perrakis, et al. (2000). Nature. 407(6805): 711-717.
Lamers, M. H., H. H. Winterwerp, et al. (2003). Embo J. 22(3): 746-756.
Lammens, A., A. Schele, et al. (2004). Curr Biol 14(19): 1778-1782.
Lammens, K., D. J. Bemeleit, et al. (2011). Cell 145(1): 54-66.
Lee, B. I., K. H. Kim, et al. (2004). Embo J 23(10): 2029-2038.
Lee, J. Y. and W. Yang (2006). Cell 127(7): 1349-1360.
Lee, S. F., M. A. Thompson, et al. (2011). Biophys J 100(7): L31-33.
Leiros, H. K., S. Kozielski-Stuhrmann, et al. (2004). J Biol Chem 279(53): 55840-55849.
Leiros, H. K. and S. M. McSweeney (2007). J Struct Biol 159(1): 92-102.
Leiros, H. K., J. Timmins, et al. (2006). Acta Crystallogr D Biol Crystallogr 62(Pt 2): 125-132.
Leiros, I., E. Moe, et al. (2005). Acta Crystallogr D Biol Crystallogr. 61(Pt 8): 1049-1056.
Leiros, I., J. Timmins, et al. (2005). EMBO J 24(5): 906-918.
Levin-Zaidman, S., J. Englander, et al. (2003). Science 299(5604): 254-256.
Lima, E., L. Wiegart, et al. (2009). Physical Review Letters 103(19).
Liu, Y., J. Zhou, et al. (2003). Proc Natl Acad Sci U S A 100(7): 4191-4196. Epub 2003 Mar 4121.
Lloyd, R. G., M. C. Porton, et al. (1988). Molecular & General Genetics 212(2): 317-324.
Lomovskaya, N., S. K. Hong, et al. (1996). J Bacteriol. 178(11): 3238-3245.
Luisi-DeLuca, C. and R. Kolodner (1994). J Mol Biol 236(1): 124-138.
MacGregor, E. A., S. Janecek, et al. (2001). Biochimica Et Biophysica Acta-Protein Structure and Molecular Enzymology 1546(1): 1-20.
Manelyte, L., C. P. Guy, et al. (2009). DNA Repair (Amst) 8(11): 1300-1310.
Matson, S. W. and J. W. George (1987). J Biol Chem 262(5): 2066-2076.
Matson, S. W. and A. B. Robertson (2006). Nucleic Acids Res 34(15): 4089-4097.
Mattimore, V. and J. R. Battista (1996). Journal of Bacteriology 178(3): 633-637.

Mazur, S. J. and L. Grossman (1991). Biochemistry. 30(18): 4432-4443.
Melby, T. E., C. N. Ciampaglio, et al. (1998). Journal of Cell Biology 142(6): 1595-1604.
Meunier-Jamin, C., U. Kapp, et al. (2004). J Biol Chem 279(24): 25830-25837.
Miao, J. W., P. Charalambous, et al. (1999). Nature 400(6742): 342-344.
Michel, B., G. Grompone, et al. (2004). Proc Natl Acad Sci U S A 101(35): 12783-12788.
Moe, E., D. R. Hall, et al. (2012). Acta Crystallographica Section D-Biological Crystallography 68: 703-712.
Moe, E., I. Leiros, et al. (2006). J Biol Chem. 281(1): 569-577.
Moreland, N., R. Ashton, et al. (2005). Acta Crystallographica Section D-Biological Crystallography 61: 1378-1385.
Morimatsu, K. and S. C. Kowalczykowski (2003). Mol Cell 11(5): 1337-1347.
Naqvi, A., E. Tinsley, et al. (2003). J Bacteriol 185(22): 6633-6639.
Oswald, C., I. B. Holland, et al. (2006). Naunyn Schmiedebergs Arch Pharmacol. 372(6): 385-399.
Pakotiprapha, D., Y. Inuzuka, et al. (2008). Mol Cell 29(1): 122-133.
Pellegrino, S., D. de Sanctis, et al. (2012). Acta Crystallogr Sect F Struct Biol Cryst Commun 68(Pt 2): 218-221.
Pellegrino, S., J. Radzimanowski, et al. (2012). Acta Crystallogr Sect F Struct Biol Cryst Commun 68(Pt 1): 81-84.
Quevillon-Cheruel, S., L. Dominique, et al. (2004). Biochimie 86(9-10): 617-623.
Reyes, E. D., P. L. Patidar, et al. (2010). Journal of Biological Chemistry 285(22): 16521-16529.
Rocha, E. P., E. Cornet, et al. (2005). PLoS Genet 1(2): e15.
Romao, C. V., E. P. Mitchell, et al. (2006). J Biol Inorg Chem 11(7): 891-902.
Saikrishnan, K., B. Powell, et al. (2009). Cell 137(5): 849-859.
Sancar, A. and W. D. Rupp (1983). Cell. 33(1): 249-260.
Sanchez, H. and J. C. Alonso (2005). Nucleic Acids Research 33(7): 2343-2350.
Sanchez, H., P. P. Cardenas, et al. (2008). Nucleic Acids Research 36(1): 110-120.
Sargentini, N. J. and K. C. Smith (1986). Mutat Res 166(1): 17-22.
Sayre, D. (2008). Acta Crystallographica Section A 64: 33-35.
Scianimanico, S., G. Schoehn, et al. (2000). Embo Journal 19(24): 6732-6741.
Segelke, B., J. Schafer, et al. (2004). Journal of Structural and Functional Genomics 5(1): 147-157.
Shen, C. H., Y. C. Chiang, et al. (2007). Mol Genet Genomics. 277(2): 149-160. Epub 2006 Dec 2005.
Singleton, M. R., M. S. Dillingham, et al. (2004). Nature 432(7014): 187-193.
Singleton, M. R., M. S. Dillingham, et al. (2007). Annu Rev Biochem 76: 23-50.
Slade, D. and M. Radman (2011). Microbiol Mol Biol Rev 75(1): 133-191.
Smith, P. C., N. Karpowich, et al. (2002). Mol Cell. 10(1): 139-149.
Sogaard, M., A. Kadziola, et al. (1993). Journal of Biological Chemistry 268(30): 22480-22484.
Soultanas, P., M. S. Dillingham, et al. (1999). Nucleic Acids Res 27(6): 1421-1428.
Sugar, F., F. Jenney, et al. (2005). Journal of Structural and Functional Genomics 6(2): 149-158.
Tanaka, M., I. Narumi, et al. (2005). J Bacteriol. 187(11): 3693-3697.
Tark, M., A. Tover, et al. (2008). DNA Repair (Amst) 7(1): 20-30.
Terwilliger, T. C. (2000). Nature Structural Biology 7: 935-939.
Thiagalingam, S. and L. Grossman (1991). J Biol Chem. 266(17): 11395-11403.
Thiagalingam, S. and L. Grossman (1993). J Biol Chem. 268(24): 18382-18389.
Timmins, J., E. Gordon, et al. (2009). Structure 17(4): 547-558.

Timmins, J., H. K. Leiros, et al. (2005). J Mol Biol. 347(5): 949-963.
Timmins, J., I. Leiros, et al. (2007). Embo J 26(13): 3260-3271.
Timmins, J., R. W. H. Ruigrok, et al. (2004). Fems Microbiology Letters 233(2): 179-186.
Timmins, J., G. Schoehn, et al. (2003a). Virology 312(2): 359-368.
Timmins, J., G. Schoehn, et al. (2003b). Journal of Molecular Biology 326(2): 493-502.
Timmins, J., S. Scianimanico, et al. (2001). Virology 283(1): 1-6.
Truglio, J. J., D. L. Croteau, et al. (2006). Chem Rev. 106(2): 233-252.
Tung, C. S., S. Joseph, et al. (2002). Nat Struct Biol. 9(10): 750-755.
Tuteja, N. and R. Tuteja (2004a). Eur J Biochem 271(10): 1835-1848.
Tuteja, N. and R. Tuteja (2004b). Eur J Biochem 271(10): 1849-1863.
Umezu, K. and R. D. Kolodner (1994). J Biol Chem 269(47): 30005-30013.
van den Bosch, M., R. T. Bree, et al. (2003). EMBO Rep 4(9): 844-849.
Van Houten, B., D. L. Croteau, et al. (2005). Mutat Res. 577(1-2): 92-117.
Veaute, X., S. Delmas, et al. (2005). EMBO J 24(1): 180-189.
Velankar, S. S., P. Soultanas, et al. (1999). Cell 97(1): 75-84.
Wang, G. and R. J. Maier (2008). Infection and Immunity 76(1): 153-160.
White, O., J. A. Eisen, et al. (1999). Science 286(5444): 1571-1577.
Wood, E. R. and S. W. Matson (1987). J Biol Chem 262(31): 15269-15276.
Yang, W. (2010). Annu Rev Biophys 39: 367-385.
Ylihonko, K., J. Tuikkanen, et al. (1996). Mol Gen Genet 251(2): 113-120.
Zahradka, K., D. Slade, et al. (2006). Nature 443(7111): 569-573.
Zaitseva, J., C. Oswald, et al. (2006). Embo J. 25(14): 3432-3443.